

CAROLINA SHENG WHEI MIAW

**DETECÇÃO DE FRAUDES EM NÉCTARES DE FRUTAS:
ANÁLISES ESPECTROSCÓPICAS ALIADAS A MÉTODOS DE
CALIBRAÇÃO E CLASSIFICAÇÃO MULTIVARIADA**

**Faculdade de Farmácia da UFMG
Belo Horizonte, MG
2018**

CAROLINA SHENG WHEI MIAW

**DETECÇÃO DE FRAUDES EM NÉCTARES DE FRUTAS: ANÁLISES
ESPECTROSCÓPICAS ALIADAS A MÉTODOS DE CALIBRAÇÃO E
CLASSIFICAÇÃO MULTIVARIADA**

Tese apresentada ao Programa de Pós-Graduação em Ciência de Alimentos da Faculdade de Farmácia da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutora em Ciência de Alimentos.

Orientadora: Profa. Dra. Scheilla Vitorino Carvalho de Souza

Coorientador: Prof. Dr. Marcelo Martins de Sena

Colaboradoras: Profa. Dra. Itziar Ruisánchez
Profa. Dra. Maria Pilar Callao

Área de concentração: Ciência de Alimentos

Linha de pesquisa: Qualidade de Alimentos

Área de Conhecimento: Avaliação e controle de Qualidade de Alimentos

**Faculdade de Farmácia da UFMG
Belo Horizonte, MG
2018**

Miaw, Carolina Sheng Whei.

M618d Detecção de fraudes em néctares de frutas: análises espectroscópicas aliadas a métodos de calibração e classificação multivariada / Carolina Sheng Whei Miaw. – 2018.

196f. : il.

Orientadora: Scheilla Vitorino Carvalho de Souza.

Coorientador: Marcelo Martins de Sena.

Colaboradoras: Itziar Ruisánchez.

Maria Pilar Callao.

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

1. Néctar - Teses. 2. Frutas - Teses. 3. Alimentos – Adulteração e inspeção – Teses. 4. Bebidas – Análise – Teses. 5. Quimometria – Teses. 6. Espectroscopia de infravermelho – Teses. 7. Validação de método – Teses. I. Souza, Scheilla Vitorino Carvalho de. II. Sena, Marcelo Martins de. III. Ruisánchez, Itziar. IV. Callao, Maria Pilar. V. Universidade Federal de Minas Gerais. Faculdade de Farmácia. VI. Título.

CDD: 664.07



UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DE ALIMENTOS

PPCCA

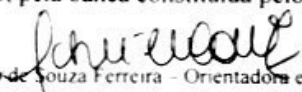
FOLHA DE APROVAÇÃO

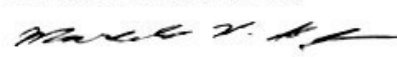
DETECÇÃO DE FRAUDES EM NÉCTARES DE FRUTAS: ANÁLISES ESPECTROSCÓPICAS ALIADAS A MÉTODOS DE CALIBRAÇÃO E CLASSIFICAÇÃO MULTIVARIADA

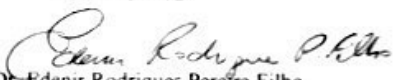
CAROLINA SHENG WHEI MIAW

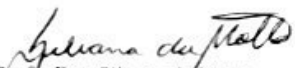
Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIA DE ALIMENTOS, como requisito para obtenção do grau de Doutor em CIÊNCIA DE ALIMENTOS, área de concentração CIÊNCIA DE ALIMENTOS.


Aprovada em 26 de novembro de 2018, pela banca constituída pelos membros:

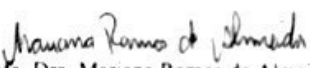

Profa. Dra. Scheilla Vitorino Carvalho de Souza Ferreira - Orientadora e Presidente da Comissão
Faculdade de Farmácia - UFMG


Prof. Dr. Marcelo Martins de Sena
ICEx - UFMG


Prof. Dr. Edenir Rodrigues Pereira Filho
UFSCAR


Profa. Dra. Silvana da Motta
Faculdade de Farmácia - UFMG


Prof. Dr. José Ailton Gonçalves
MAPA


Profa. Dra. Mariana Ramos de Almeida
ICEx - UFMG

Belo Horizonte, 26 de novembro de 2018.

AGRADECIMENTOS

Aos meus orientadores Profs. Scheilla Vitorino Carvalho de Souza e Marcelo Martins de Sena pelos ensinamentos, paciência, dedicação, amizade e confiança.

Aos membros da Comissão Examinadora Profs. Edenir Rodrigues Pereira Filho, José Ailton Gonçalves, Mariana Ramos de Almeida e Silvana da Motta pelas avaliações e contribuições para o trabalho.

À equipe do Laboratório de Bromatologia – Unidade de Pesquisa Análise de Alimentos (BRO-UPAA) do Departamento de Alimentos (ALM) da Faculdade de Farmácia (FAFAR) da UFMG pelo apoio e amizade, com dedicatória especial aos que me auxiliaram na execução do projeto: Alessandro Rangel Carolino Sales Silva, Maria Luisa Cunha, Pedro Paulo Borges Santos e Ronália Leite Alvarenga.

Ao Grupo de Quimiometria Aplicada à Química Analítica e Técnicas Espectroscópicas (GQQATE) do Departamento de Química (DQ) da UFMG, em especial, à Camila Assis pela parceria no estudo de calibração multivariada e à Profa. Poliana Macedo dos Santos pela assistência e parceria em relação às análises por ressonância magnética nuclear.

À Profa. Silvana da Motta pelos ensinamentos relativos à fabricação dos néctares e empréstimo de equipamentos do Laboratório de Operações e Tecnologia do ALM/FAFAR/UFMG.

À Profa. Adriana Silva França pelo empréstimo do espectrofotômetro do Laboratório de Biocombustíveis da Escola de Engenharia da UFMG e às MSc Verônica Belchior e Andreia Suzuki pela assistência com sua utilização.

À Profa. Clésia Cristina Nascentes e ao Programa Ciências Forenses (“Pró-Forenses” - Edital CAPES nº. 25/2014), pelo empréstimo do espectrofotômetro do Departamento de Química e ao MSc Adão Edson de Barcelos Júnior pela ajuda em seu uso.

Aos servidores e técnicos-administrativos da FAFAR/UFMG pela presteza e cordialidade, em especial à Edna Aparecida de Souza, Marcos da Costa Lage, Maria Paula, Úrsula Regiane Martins Rodrigues, Gabriel Cassani e Gabriel Rotsen.

Ao pesquisador Giuliano Elias Pereira da Embrapa Uva & Vinho pelo envio de amostras de uva Isabel de Petrolina, Pernambuco.

Às Profas. Itziar Ruisánchez e Maria Pilar Callao do *Chemometrics, Qualimetrics and Nanosensors Group (Department of Analytical Chemistry and Organic Chemistry)* pelo aprendizado adquirido durante o doutorado sanduíche na *Universitat Rovira i Virgili*.

Ao Prof. Paolo Oliveri da *Università degli Studi di Genova* pelo envio da rotina desenvolvida para o método de mínimos quadrados parciais por modelagem de densidade (PLS-DM) e assistência relacionada ao seu uso.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro na concessão das bolsas dos Programas Demanda Social (Código de Financiamento 001) e de Doutorado Sanduíche no Exterior (nº 88881.132172/2016-01).

Aos funcionários Aline Gozzi e Nilson César Castanheira Guimarães do Laboratório Nacional Agropecuário (LANAGRO-MG) pela parceria nos experimentos utilizando ressonância magnética nuclear no domínio do tempo (RMN-DT) e aos funcionários Carla Amaral, Emanuel Novaes Vasconcelos, Glauco Antônio Teixeira e Lidiane Leal Duarte Lisboa pelo auxílio nos estudos iniciais do projeto.

Ao meu marido Alexandre, por sempre me incentivar, apoiar e ser meu companheiro durante todos esses anos. À minha família pelo apoio incondicional e aos amigos pelos momentos de diversão.

A Deus pela vida e bênçãos diárias!

SUMÁRIO

1.	INTRODUÇÃO.....	19
2.	OBJETIVOS.....	22
2.1	OBJETIVO GERAL.....	22
2.2	OBJETIVOS ESPECÍFICOS.....	22
3.	REVISÃO DE LITERATURA.....	25
3.1	NÉCTARES DE FRUTAS.....	25
3.1.1	Aspectos nutricionais.....	25
3.1.2	Aspectos econômicos.....	26
3.1.3	Legislação.....	27
3.1.4	Métodos para detecção de fraudes em néctares.....	29
3.2	TÉCNICAS ESPECTROSCÓPICAS.....	32
3.2.1	Espectroscopia ATR-FTIR.....	32
3.2.2	Espectroscopia RMN-DT.....	33
3.3	QUIMIOMETRIA.....	36
3.3.1	Calibração multivariada.....	36
3.3.1.1	Regressão por Mínimos Quadrados Parciais (PLS).....	37
3.3.2	Classificação multivariada.....	40
3.3.2.1	Modelagem independente e flexível por analogia de classe (SIMCA)..	42
3.3.2.2	Análise discriminante por mínimos quadrados parciais (PLS-DA).....	44
3.3.2.3	Mínimos quadrados parciais por modelagem de densidade (PLS-DM)	45
3.3.2.4	Mínimos quadrados parciais de uma classe (OCPLS).....	46
3.3.2.5	SIMCA orientado aos dados (DD-SIMCA).....	48
3.3.2.6	Algoritmo de Kennard-Stone.....	50
3.3.2.7	Validação cruzada.....	51
3.3.4	Seleção de variáveis.....	52
3.3.4.1	Regressão de mínimos quadrados parciais por intervalo (iPLS).....	53
3.3.4.2	Seleção dos preditores ordenados (OPS).....	53
3.3.4.3	Algoritmo genético (GA)	55
3.3.4.4	VIP escores.....	57
3.3.4.5	iPLS combinado a outro método de seleção de variáveis.....	58
3.4	VALIDAÇÃO ANALÍTICA MULTIVARIADA.....	59
3.4.1	Abordagem quantitativa.....	59
3.4.2	Abordagem qualitativa.....	62
4.	MATERIAL E MÉTODOS.....	64
4.1	EQUIPAMENTOS.....	64
4.2	MATERIAL.....	65
4.2.1	Determinação de acidez total titulável e pH.....	65
4.2.2	Determinação de açúcares totais.....	65
4.2.3	Determinação do teor de ácido ascórbico.....	66
4.2.4	Análise por espectroscopia ATR-FTIR.....	66
4.3	REAGENTES, PADRÕES E OUTROS.....	66
4.4	AMOSTRAS.....	68
4.4.1	Frutas.....	68
4.4.2	Néctares comerciais.....	71
4.4.3	Néctares formulados.....	73
4.5	MÉTODOS DE ANÁLISE.....	77
4.5.1	Determinação dos parâmetros do PIQ.....	77

4.5.2	Análise por espectroscopia ATR-FTIR.....	77
4.5.3	Análise quimiométrica.....	78
4.5.4	Seleção de variáveis.....	78
4.5.5	Análise por RMN-DT.....	79
	CAPÍTULO I. DETERMINAÇÃO DAS CONCENTRAÇÕES DE FRUTAS PRINCIPAIS EM NÉCTARES ADULTERADOS POR ESPECTROSCOPIA POR REFLETÂNCIA TOTAL ATENUADA NO INFRAVERMELHO MÉDIO COM TRANSFORMADA DE FOURIER (ATR-FTIR) COMBINADA COM MÉTODOS DE CALIBRAÇÃO MULTIVARIADA E SELEÇÃO DE VARIÁVEIS	80
I.1	PRODUÇÃO.....	82
I.2	INTRODUÇÃO.....	83
I.3	MATERIAL E MÉTODOS.....	85
I.3.1	DELINEAMENTO EXPERIMENTAL.....	85
I.3.1.1	Diferenças varietais.....	85
I.3.1.2	PIQ dos néctares comerciais.....	86
I.3.1.3	Néctares formulados.....	86
I.3.1.4	Avaliação da conformidade dos néctares formulados.....	90
I.3.1.5	Validação quantitativa.....	90
I.4	RESULTADOS E DISCUSSÃO.....	92
I.4.1	DIFERENÇAS VARIETAIS.....	92
I.4.2	CARACTERIZAÇÃO FÍSICO-QUÍMICA DOS NÉCTARES COMERCIAIS.....	93
I.4.3	CARACTERIZAÇÃO FÍSICO-QUÍMICA DOS NÉCTARES FORMULADOS.....	95
I.4.4	MODELO DE CALIBRAÇÃO MULTIVARIADA PARA QUANTIFICAÇÃO DAS FRUTAS PRINCIPAIS EM NÉCTARES NÃO ADULTERADOS E ADULTERADOS.....	97
I.4.4.1	Atribuição espectral dos néctares formulados.....	100
I.4.4.2	Pré-processamento dos dados.....	101
I.4.4.3	Modelos de calibração multivariada.....	101
I.4.4.4	Validação dos modelos iPLS-OPS para néctares de laranja, uva, pêsego e maracujá.....	107
I.5	CONCLUSÃO.....	113
	CAPÍTULO II. DETECÇÃO DE ADULTERANTES EM NÉCTARES DE UVA POR ESPECTROSCOPIA POR REFLETÂNCIA TOTAL ATENUADA NO INFRAVERMELHO MÉDIO COM TRANSFORMADA DE FOURIER (ATR-FTIR) E ESTRATÉGIAS DE CLASSIFICAÇÃO MULTIVARIADA.....	115
II.1	PRODUÇÃO.....	116
II.2	INTRODUÇÃO.....	117
II.3	MATERIAL E MÉTODOS.....	120
II.3.1	MÉTODOS.....	120
II.3.1.1	Análise dos dados multivariados.....	120
II.3.2	DELINEAMENTO.....	120
II.3.2.1	Formulação dos néctares.....	120
II.3.2.2	Parâmetros de desempenho.....	122
II.4	RESULTADOS E DISCUSSÃO.....	123
II.4.1	ANÁLISE ESPECTRAL.....	123
II.4.2	PRÉ-PROCESSAMENTO DOS DADOS.....	124
II.4.3	ANÁLISE DE COMPONENTES PRINCIPAIS (PCA).....	124
II.4.4	MODELOS MULTICLASSES.....	125

II.4.5	MODELOS DE CLASSE ÚNICA.....	130
II.5	CONCLUSÃO.....	132
CAPÍTULO III. SELEÇÃO DE VARIÁVEIS ALIADA À CLASSIFICAÇÃO MULTIVARIADA COM O OBJETIVO DE DETECTAR ADULTERANTES INDIVIDUAIS E SUAS MISTURAS EM NÉCTARES DE UVA.....		134
III.1	PRODUÇÃO.....	135
III.2	INTRODUÇÃO.....	136
III.3	MATERIAL E MÉTODOS.....	140
III.3.1	DELINEAMENTO.....	140
III.3.1.1	Formulação dos néctares.....	140
III.3.1.2	Parâmetros de desempenho.....	142
III.4	RESULTADOS E DISCUSSÃO.....	143
III.4.1	ANÁLISE ESPECTRAL.....	143
III.4.2	PRÉ-PROCESSAMENTO DOS DADOS.....	144
III.4.3	PREDIÇÕES COM ESPECTRO COMPLETO.....	144
III.4.4	PREDIÇÕES COM APLICAÇÃO DOS MÉTODOS DE SELEÇÃO DE VARIÁVEIS.....	146
III.5	CONCLUSÃO.....	155
CAPÍTULO IV. COMPARAÇÃO DE DIFERENTES MÉTODOS DE CLASSIFICAÇÃO MULTIVARIADA PARA DETECÇÃO DE ADULTERAÇÕES EM NÉCTARES DE UVA POR RESSONÂNCIA MAGNÉTICA NUCLEAR NO DOMÍNIO DE TEMPO (RMN-DT)		156
IV.1	PRODUÇÃO.....	157
IV.2	INTRODUÇÃO.....	158
IV.3	MATERIAL E MÉTODOS.....	160
IV.3.1	MÉTODOS.....	160
IV.3.1.1	Análise de dados.....	160
IV.3.1.2	Construção dos modelos de classificação.....	160
IV.3.2	DELINEAMENTO.....	161
IV.3.2.1	Formulação dos néctares.....	161
IV.3.2.2	Parâmetros de desempenho.....	162
IV.4	RESULTADOS E DISCUSSÃO.....	163
IV.4.1	RESULTADOS DE RELAXAÇÃO DE T ²	163
IV.4.2	MODELO PCA.....	165
IV.4.3	MÉTODOS DE CLASSE ÚNICA.....	165
IV.4.4	MODELOS MULTICLASSES.....	171
IV.5	CONCLUSÃO.....	177
5.	CONSIDERAÇÕES FINAIS.....	178
REFERÊNCIAS BIBLIOGRÁFICAS.....		180

LISTA DE TABELAS

Tabela 1 Padrões de Identidade e Qualidade (PIQ) dos néctares de caju, maracujá, pêssego, maçã, laranja e uva.....	28
Tabela 2 Técnicas analíticas para análise de fraudes em bebidas à base de frutas.....	30
Tabela 3 Descrição das frutas utilizadas no preparo dos néctares com respectivas variedades, cor, grau de maturação condições de refrigeração, branqueamento e rendimento.....	69
Tabela 4 Datas de validade dos néctares comerciais adquiridos.....	72
Tabela I.1 Parâmetros de desempenho calculados na validação quantitativa.....	91
Tabela I.2 Valores máximo (MAX) e mínimo (MIN), médias (MED) e desvio padrão (DP) obtidos para os parâmetros sólidos solúveis, açúcares totais, acidez total titulável e teor de ácido ascórbico dos néctares comerciais.....	94
Tabela I.3 Discriminação das médias pelo teste de Tukey dos néctares formulados para os parâmetros sólidos solúveis, açúcares totais, acidez total titulável e teor de ácido ascórbico.....	96
Tabela I.4 Comparação dos parâmetros estatísticos estimados para os modelos com espectro completo e com métodos de seleção de variáveis.....	105
Tabela I.5 Figuras de mérito dos modelos obtidos com iPLS-OPS para néctares de uva, laranja, pêssego e maracujá.....	108
Tabela II.1 Previsões de classe dos modelos SIMCA e PLS-DA das amostras dos conjuntos treinamento e teste.....	128
Tabela II.2 Sensibilidade e seletividade dos modelos SIMCA, PLS-DA e PLS-DM utilizando a estratégia de classe única.....	131

Tabela III.1 Predições dos métodos SIMCA e PLS-DA utilizando o espectro completo.....	145
Tabela III.2 Resultados do modelo SIMCA combinado com os métodos de seleção de variáveis iPLS, GA e VIP escores.....	149
Tabela III.3 Resultados do modelo PLS-DA combinado com os métodos de seleção de variáveis iPLS, GA e VIP escores.....	151
Tabela III.4 Resultados do modelo PLS-DA combinado com os métodos de seleção de variáveis iPLS-GA e iPLS-VIP.....	153
Tabela III.5 Sensibilidade (SEN) e seletividade (SEL) para os modelos PLS-DA com espectro completo e com os diferentes métodos de seleção de variáveis.....	154
Tabela IV.1 Parâmetros de desempenho estimados a partir dos conjuntos teste das amostras não adulterados e todas as amostras adulteradas para os modelos de classe única SIMCA, OCPLS e DD-SIMCA.....	170
Tabela IV.2 Predições multiclases de SIMCA e PLS-DA das amostras da classe não adulterada (UN), adulterada com caju (CAS), adulterada com maçã (APP) para conjuntos de treinamento e teste e das amostras adulteradas com caju e maçã (CAS+APP).....	172
Tabela IV.3 Parâmetros de desempenho sensibilidade (SEN), seletividade (SEL), taxa de inconclusivos (IR) e eficiência (EFF) para modelos multiclases SIMCA e PLS-DA.....	174

LISTA DE FIGURAS

Figura 1 Decomposição dos dados pelo método dos mínimos quadrados parciais (PLS).....	39
Figura 2 Representação esquemática do método seleção dos preditores ordenados (OPS).....	55
Figura 3 Fluxograma básico do processamento dos néctares.....	73
Figura I.1 Delineamento do experimento de diferenças varietais.....	86
Figura I.2 Delineamento do estudo de adulterações com xarope, maçã e caju em diferentes níveis.....	88
Figura I.3 Gráfico de valores de T^2 de <i>Hotelling versus</i> resíduos Q para detecção de <i>outliers</i> no modelo PCA construído com néctares de maçã.....	92
Figura I.4 Escores em PC1 e PC2 para o modelo construído com espectros de néctares de maçã.....	93
Figura I.5 Espectros representativos dos néctares de uva (1), laranja (2), pêssigo (3) e maracujá (4), não adulterado (a), adulterado pela diluição com xarope (b), adição de suco de maçã (c) e adição de suco de caju (d), nos níveis mais altos.....	99
Figura I.6 Espectros completos das amostras formuladas de néctares de uva (a), laranja (b), pêssigo (c) e maracujá (d) com as variáveis destacadas selecionadas pelo iPLS-OPS.....	103
Figura I.7 Valores de referência versus valores preditos para o modelo iPLS-OPS de uva (a), laranja (b), pêssigo (c) e maracujá (d).....	109
Figura II.1 Esquema da formulação das amostras de néctares de uva.....	121

Figura II.2 Espectros médios pré-processados da classe não adulterada (linha tracejada azul), classe adulterada com caju (linha sólida verde) e classe adulterada com maçã (linha traço-ponto vermelha).....	124
Figura II.3 Escores de PC1 <i>versus</i> PC2 das amostras não-adulteradas (triângulo vermelho), adulterado com caju (círculo verde) e adulterado com maçã (quadrado azul).....	125
Figura II.4 Previsões PLS-DA para cada classe: a) não adulterada (vermelho), b) adulterada com caju (verde) e c) adulterada com maçã (azul).....	127
Figura III.1 Visão geral das etapas para desenvolvimento do método de classificação multivariada com seleção de variáveis.....	139
Figura III.2 Delineamento experimental para a formulação dos néctares.....	141
Figura III.3 Espectros de todas as amostras de néctares de uva analisadas após pré-processamento com correção do espalhamento multiplicativo.....	143
Figura III.4 Gráfico das variáveis selecionadas por a) iPLS, com valores de RMSECV de cada intervalo, b) GA e c) VIP escores obtidos para as três classes.....	147
Figura IV.1 Delineamento experimental da formulação dos néctares de uva.....	162
Figura IV.2 a) Curvas de relaxação obtidas com a normalização das médias CPMG para as amostras não adulteradas (linha azul pontilhada), adulteradas com caju (linha verde sólida) e adulteradas com maçã (linha vermelha tracejada). b) Distribuição dos tempos de relaxação T2 estimados pela transformada inversa de Laplace das curvas CPMG obtidas para amostra não adulterada (linha azul pontilhada), adulteradas com caju (linha verde sólida) e adulteradas com maçã (linha vermelha tracejada).....	164
Figura IV.3 Análise de componentes principais dos dados de relaxação CPMG das amostras de néctares de uva não adulteradas (triângulos vermelhos), adulteradas com caju (estrelas verdes) e adulteradas com maçã (quadrados azuis).....	165

Figura IV.4 Gráficos de aceitação obtidos com DD-SIMCA para os conjuntos treinamento (a) e teste (b). h_i e v_i são, respectivamente, os valores da distância da pontuação e da distância ortogonal para a amostra $i = 1, \dots, l$167

Figura IV.5 Gráficos obtidos pelo modelo OCPLS das distâncias do scores (SD) versus resíduos absolutos centralizados (ACR) das amostras do conjunto treinamento (a) e teste (b).....169

LISTA DE ACRÔNIMOS

ABIR	Associação Brasileira das Indústrias de Refrigerantes e de Bebidas Não Alcoólicas
APP	Classe adulterada com maçã
ATR	Reflectância total atenuada (<i>Attenuated total reflectance</i>)
ATR-FTIR	Refletância total atenuada no infravermelho médio com transformada de Fourier (<i>Attenuated total reflectance Fourier-transform infrared</i>)
CAS	Classe adulterada com caju
CAS+APP	Amostras adulteradas com caju e maçã
DD-SIMCA	SIMCA orientado aos dados (<i>Data driven SIMCA</i>)
DPR _r	Desvio padrão relativo no nível de repetibilidade
DPR _R	Desvio padrão relativo no nível de precisão intermediária
ER	Erro relativo
FT-IR	Infravermelho com transformada de Fourier (<i>Fourier transform infrared</i>)
FOM	Figuras de mérito
GA	Algoritmo genético (<i>Genetic Algorithm</i>)
IN	Instrução normativa
IR	Resultados inconclusivos
iPLS	Regressão de mínimos quadrados parciais por intervalo (<i>Interval partial least squares</i>)

IV	Infravermelho
LOD	Limite de detecção
MA	Múltiplas atribuições
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
MSC	Correção do espalhamento multiplicativo (<i>Multiplicative scatter correction</i>)
NA	Não atribuído
n_{VL}	Número de variáveis latentes
OCPLS	Mínimos quadrados parciais de uma classe (<i>One-class partial least squares</i>)
OPS	Seleção dos preditores ordenados (<i>Ordered predictors selection</i>)
PCA	Análise de componentes principais (<i>Principal component analysis</i>)
PIQ	Padrão de identidade e qualidade
PLS	Mínimos quadrados parciais (<i>Partial least squares</i>)
PLS-DA	Análise discriminante por mínimos quadrados parciais (<i>Partial least squares discriminant analysis</i>)
PLS-DM	Mínimos quadrados parciais por modelagem de densidade (<i>Partial least squares density modeling</i>)
r	Coefficiente de correlação
R_c	Coefficiente de correlação de calibração
R_{cv}	Coefficiente de correlação de validação cruzada
R_p	Coefficiente de correlação de predição

RMN-DT	Ressonância magnética nuclear no domínio do tempo
RMSEC	Raiz quadrada dos erros médios quadráticos de calibração (<i>root mean square error of calibration</i>)
RMSECV	Raiz quadrada dos erros médios quadráticos de validação cruzada (<i>root mean square error of cross-validation</i>)
RMSEP	Raiz quadrada dos erros médios quadráticos de previsão (<i>root mean square error of prediction</i>)
RPD	Relação de desempenho do desvio
SEL	Seletividade
SEN	Sensibilidade
SIMCA	Modelagem independente e flexível por analogia de classe (<i>Soft independent modeling of class analogy</i>)
VIP Scores	Variáveis importantes na projeção dos escores (<i>Variable importance in projection scores</i>)
VLs	Variáveis latentes
UN	Classe não adulterada

RESUMO

Devido ao crescente avanço no consumo de bebidas à base de frutas, os néctares de frutas se destacam, porque o preço final destas bebidas é bem menor do que os sucos integrais pasteurizados. O néctar de laranja é o mais consumido mundialmente e os néctares de uva, pêsego, laranja e maracujá, nesta ordem, são os mais consumidos no Brasil. Existem muitos tipos conhecidos de fraudes em sucos e néctares que ocorrem durante o processo de fabricação, sendo frequentes a diluição com água, a adição de xarope de açúcar e a adição não declarada de diferentes espécies. Técnicas espectroscópicas podem ser utilizadas para avaliar a estrutura e a composição dos alimentos. A combinação destas técnicas com métodos quimiométricos de análise é muito utilizada para determinação de autenticidade em alimentos. Os principais objetivos deste trabalho foram: a) análise e comparação do perfil espectral das variedades de maçã Gala e Fuji envolvidas na produção de néctares no Brasil; b) caracterização dos néctares comerciais (uva, pêsego, laranja e maracujá) em relação aos seus Padrões de Identidade e Qualidade; c) desenvolvimento e validação de modelos de calibração multivariada para quantificação do teor de fruta principal em néctares de uva, pêsego, laranja e maracujá não adulterados e adulterados com xarope, suco de caju e suco de maçã; d) desenvolvimento e validação de modelos de classificação supervisionada para discriminação e classificação dos adulterantes caju e maçã, isoladamente, em néctares de uva utilizando espectroscopia ATR-FTIR (refletância total atenuada no infravermelho médio com transformada de Fourier); e) desenvolvimento e validação de modelos de classificação supervisionada para discriminação e classificação dos adulterantes caju e maçã, em misturas, em néctares de uva utilizando ATR-FTIR; f) desenvolvimento e validação de modelos de classificação supervisionada para discriminação e classificação dos adulterantes caju e maçã, isoladamente e em misturas, em néctares de uva, utilizando ressonância magnética nuclear no domínio do tempo (RMN-DT). Os métodos e modelos desenvolvidos e validados foram adequados aos seus propósitos de uso, com potencial contribuição na detecção de fraudes em néctares de frutas.

Palavras-chave: Néctares de frutas. Adulteração em bebidas. ATR-FTIR. RMN-DT. Quimiometria. Métodos de classificação multivariada.

ABSTRACT

Due to the increasing advancement in the consumption of fruit drinks, fruit nectars stand out because their final price is much lower than pasteurized whole juices. Orange nectar is the most consumed worldwide, and the most consumed nectars in Brazil are grape, peach, orange and passion fruit, in this order. There are many known types of adulteration in juices and nectars that occur during the manufacturing process and some of them are frequent, such as: dilution with water, addition of sugar syrup and undeclared addition of different species. Spectroscopic techniques can be used to evaluate the structure and composition of food and beverages. The combination of these techniques with chemometric methods is widely used to determine authenticity in food. The main objectives of this work were: a) analysis and comparison of the spectral profile of Gala and Fuji apple varieties involved in the production of nectars in Brazil; b) characterization of commercial nectars (grape, peach, orange and passion fruit) in relation to their Identity and Quality Standards; c) development and validation of multivariate calibration models for quantification of the main fruit content in nectars (grape, peach, orange and passion fruit) unadulterated and adulterated with syrup, cashew juice and apple juice; d) development and validation of supervised classification models for discrimination and classification of cashew and apple adulterants, isolated, in grape nectars using attenuated total reflectance Fourier-transform infrared (ATR-FTIR) spectroscopy; e) development and validation of supervised classification models for discrimination and classification of cashew and apple adulterants, as a mixture, in grape nectars using ATR-FTIR; f) development and validation of supervised classification models for discrimination and classification of cashew and apple adulterants, isolated and as a mixture, in grape nectars using time-domain nuclear magnetic resonance (TD-NMR). The methods and models developed and validated were adequate for their purposes of use, with potential contribution for fraud detection in fruit nectars.

Key-words: Fruit nectars. Adulteration in beverages. ATR-FTIR. Multivariate calibration. Chemometrics. Multivariate classification methods.

1 INTRODUÇÃO

O crescente avanço no consumo de bebidas à base de frutas é uma realidade e se deve ao aumento da preocupação com a saúde, levando a busca por bebidas com características nutricionais que auxiliem na prevenção e controle de doenças. A substituição do consumo de refrigerantes por bebidas mais saudáveis e com apelo à praticidade é uma tendência mundial, aumentando a demanda por produtos naturais, orgânicos e enriquecidos nutricionalmente (Esperancini, 2005).

Néctar, por definição, é a bebida não fermentada, obtida da diluição em água potável da parte comestível do vegetal ou de seu extrato, adicionada de açúcares e destinada ao consumo direto (Brasil, 2009).

De acordo com a Associação Brasileira das Indústrias de Refrigerantes e de Bebidas Não Alcoólicas (ABIR), dentre todas as categorias de bebidas não alcoólicas, as quais incluem refrigerantes, água mineral, néctares, sucos concentrados, energéticos, isotônicos, chás prontos para beber, refrescos prontos para beber, refrescos em pó e bebidas a base de soja, o terceiro maior consumo em 2013 foi o de néctares (ABIR, 2013).

No Brasil, os néctares mais consumidos são: uva, pêssego, laranja, maracujá, goiaba, caju e manga, nesta ordem (ABIR, 2015). A legislação brasileira estabelece um padrão de identidade e qualidade (PIQ) para cada néctar de fruta, incluindo os percentuais mínimos de polpa, sendo 50 % para uva e laranja, 40 % para pêssego e 10 % para maracujá. Para algumas frutas também são determinados limites para os parâmetros sólidos solúveis (°Brix), acidez total titulável, açúcares totais e teor de ácido ascórbico (MAPA, 2003; 2013).

As bebidas à base de frutas são produtos muito susceptíveis à adulteração (quanto maior for o grau de diluição do produto, maiores as chances de adulteração), principalmente devido à sua composição química e à ampla variação natural das frutas. Os tipos mais comuns de fraude incluem a adição de água ou xarope, acidificação, adição ou substituição por frutas mais baratas e adição de corantes ou flavorizantes (Singhal, Kulkarni & Rege, 1997; Asadpoor, Ansarin & Nemati, 2014; Rinke, 2016). Esse fato se soma aos poucos requisitos estabelecidos

do PIQ para alguns néctares (como laranja e uva, em que apenas o mínimo conteúdo de fruta é estabelecido), contribuindo para a atuação de indústrias visando ganho financeiro através de práticas fraudulentas (European Union, 2013).

Devido às suas características sensoriais e ao seu menor custo, o suco de maçã é comumente usado como adulterante. Outra fruta que também pode ser empregada como substituto é o caju, devido as suas características sensoriais. Uma vez que a comercialização de néctares mistos (contendo mais de uma fruta) está em expansão no mercado, para garantir sua autenticidade e para evitar a fraude, a presença declarada de cada fruta deve ser confirmada (Soria *et al.*, 2008).

Diferentes técnicas analíticas têm sido reportadas recentemente na literatura com o objetivo de avaliar e detectar vários tipos de adulteração em bebidas à base de frutas tais como cromatografia líquida de alta eficiência (HPLC) (Han *et al.*, 2012; Abad-García *et al.*, 2014; Asadpoor, Ansarin & Nemat, 2014), cromatografia gasosa acoplada à espectrometria de massas (GC-MS) (Nuncio-Jáuregui *et al.*, 2014) e reação em cadeia da polimerase (PCR) (Han *et al.*, 2012; Pardo, 2015). No entanto, essas técnicas são laboriosas, lentas e relativamente caras, envolvendo, ainda, consumo elevado de solventes e gerando resíduos.

A aplicação de técnicas espectroscópicas como a espectroscopia no ultravioleta-visível (UV-VIS) (Boggia *et al.*, 2013), espectrofluorimetria (Ammari, Redjda & Rutledge, 2015) e espectroscopia no infravermelho médio (He, Rodriguez-Saona & Giusti, 2007; Shah *et al.*, 2010; Snyder *et al.*, 2014; Ellis *et al.*, 2016; Shen *et al.*, 2016) atende à tendência na análise de alimentos por métodos mais simples, não destrutivos, ambientalmente amigáveis e com respostas rápidas (Fernández-González *et al.*, 2014; Kirtil *et al.*, 2017).

Contudo, os trabalhos citados anteriormente avaliaram diferentes tipos de fraudes na matriz suco e estavam restritos aos sabores romã, laranja e uva, sendo que em nenhum estudo foi avaliada a matriz néctar e nem com os adulterantes maçã e caju.

Neste trabalho serão empregadas a espectroscopia por refletância total atenuada no infravermelho médio com transformada de Fourier (ATR-FTIR) e a

ressonância magnética nuclear no domínio do tempo (RMN-DT) combinadas com ferramentas quimiométricas, objetivando extrair o máximo de informação possível sobre dados coletados.

Modelos de calibração multivariada são aplicados à análise química do alimento de forma a desenvolver métodos que estimem propriedades de interesse da amostra baseados nas medidas de suas propriedades químicas enquanto métodos de classificação atribuem uma determinada amostra a uma ou mais classes baseados nas medidas realizadas para caracterizá-la (Bro, 2003; Bevilacqua *et al.*, 2013).

Assim, para aplicação da calibração multivariada optou-se por utilizar o método de mínimos quadrados parciais (PLS) e para a aplicação da classificação multivariada foram estudados os métodos de modelagem independente e flexível por analogia de classe (SIMCA), SIMCA orientado aos dados (DD-SIMCA), análise discriminante por mínimos quadrados parciais (PLS-DA), mínimos quadrados parciais por modelagem de densidade (PLS-DM) e regressão de mínimos quadrados parciais de uma classe (OCPLS).

Métodos de seleção de variáveis permitem eliminar variáveis consideradas irrelevantes, ruidosas ou com pouca influência preditiva melhorando as previsões e/ou reduzindo a complexidade do modelo (Andersen & Bro, 2010). Os métodos de seleção de variáveis abordados neste trabalho foram a regressão de mínimos quadrados parciais por intervalo (iPLS), a seleção dos preditores ordenados (OPS), o algoritmo genético (GA) e as variáveis importantes na projeção dos escores (VIP escores).

Desta forma, na presente tese foram empregadas diferentes técnicas espectroscópicas (ATR-FTIR e RMN-DT) aliadas a ferramentas quimiométricas (métodos de calibração e de classificação multivariada) e a métodos de seleção de variáveis para a análise de néctares de frutas adulterados por diluição com xarope e por substituição da fruta principal por suco de maçã e suco de caju, individualmente ou como mistura, de forma a quantificar a concentração da fruta principal e classificar as adulterações de acordo com o tipo.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Desenvolver e validar métodos para detecção de fraudes em néctares de frutas por diluição com xarope e por substituição com sucos de maçã e caju, empregando a espectroscopia ATR-FTIR combinada a métodos de calibração e classificação multivariada e a RMN-DT combinada com métodos de classificação multivariada.

2.2 OBJETIVOS ESPECÍFICOS

- Analisar e comparar o perfil espectral das diferentes variedades de maçã envolvidas na produção de néctares no Brasil, Gala e Fuji.
- Caracterizar diferentes marcas comerciais de néctares de uva, pêssego, laranja, maracujá, caju e maçã em relação aos respectivos parâmetros do PIQ.
- Desenvolver e validar modelos de calibração multivariada PLS aliados a métodos de seleção de variáveis para quantificar o teor de fruta principal em néctares de uva, pêssego, laranja e maracujá não adulterados e adulterados com xarope, suco de caju e suco de maçã:
 - formular néctares de uva, pêssego, laranja e maracujá, simulando adulteração por diluição com xarope, por adição de suco de maçã e suco de caju, e caracterizá-los de acordo com os respectivos parâmetros do PIQ;
 - analisar os néctares formulados por espectroscopia ATR-FTIR;

- construir e validar modelos de calibração multivariada PLS utilizando os métodos de seleção de variáveis iPLS, OPS, GA, iPLS-OPS e iPLS-GA e selecionando o de melhor desempenho analítico.
- Desenvolver e validar modelos de classificação supervisionada para discriminar e classificar os adulterantes caju e maçã, isoladamente, em néctar de uva, empregando ATR-FTIR:
 - formular néctares de uva, simulando adulteração por adição de suco de maçã e de suco de caju, isoladamente;
 - analisar os néctares formulados por espectroscopia ATR-FTIR;
 - construir e validar modelos de classificação supervisionada empregando SIMCA, PLS-DA e PLS-DM em estratégia de classe única (*one-class*) e SIMCA e PLS-DA em estratégia multiclases;
 - comparar o desempenho dos métodos utilizados.
- Desenvolver e validar modelos de classificação supervisionada para discriminar e classificar os adulterantes caju e maçã, isoladamente e em misturas, em néctar de uva, empregando ATR-FTIR:
 - formular néctares de uva, simulando adulteração por adição de suco de maçã, de suco de caju e de misturas de suco de maçã e caju;
 - analisar os néctares formulados por espectroscopia ATR-FTIR;
 - construir e validar modelos de classificação supervisionada empregando SIMCA e PLS-DA em estratégia multiclases;
 - utilizar métodos de seleção de variáveis iPLS, GA, VIP escores e combinações iPLS-GA e iPLS-VIP de forma a melhorar a interpretação do modelo e sua capacidade preditiva;
 - comparar o desempenho dos métodos utilizados.

- Desenvolver e validar modelos de classificação supervisionada para discriminar e classificar os adulterantes caju e maçã, isoladamente e em misturas, em néctar de uva, utilizando RMN-DT:
 - analisar os néctares formulados por espectroscopia RMN-DT;
 - construir e validar modelos de classificação supervisionada empregando SIMCA, OCPLS e DD-SIMCA em estratégia de classe única e SIMCA e PLS-DA em estratégia multiclases;
 - comparar o desempenho dos métodos utilizados.

3 REVISÃO DE LITERATURA

3.1 NÉCTARES DE FRUTAS

3.1.1 Aspectos nutricionais

A indústria de sucos e néctares de frutas é um dos negócios agrícolas mais importantes do mundo. Apesar das frutas serem itens perecíveis e de colheita sazonal, seu consumo pode ser disponibilizado continuamente devido às tecnologias empregadas no processamento de sucos e néctares (Asadpoor, Ansarin & Nemat, 2014).

Os sucos de frutas possuem uma grande variedade de nutrientes em sua composição, tais como vitaminas, minerais e carboidratos. Apesar de conter proteínas e lipídeos, estes estão presentes em pequenas quantidades. Componentes como cálcio, ferro, vitamina A, tiamina (vitamina B1), riboflavina (vitamina B2) e ácido ascórbico (vitamina C) são importantes nutrientes, encontrados em maiores concentrações nos sucos de frutas do que em outros tipos de alimentos. Além disso, algumas frutas contêm compostos que apresentam atividade antioxidante, tais como tocoferol (vitamina E), beta-caroteno e flavonoides (Ashurst, 2005).

O valor nutricional das bebidas à base de frutas também está relacionado ao tipo de fruta utilizado, ao processamento e ao grau de diluição empregado. O teor de vitaminas, por exemplo, é menor quando comparado à fruta *in natura* e as perdas dependem do tipo de fruto. No preparo dos néctares, parte da fibra da fruta é removida e o valor calórico é aumentado devido à adição de açúcar (Guiné, 2016).

Devido ao aumento da preocupação com a saúde, bebidas à base de frutas tiveram seu consumo aumentado nos últimos anos (ABIR, 2014). Os consumidores estão em busca de bebidas com características nutricionais que possam auxiliar na

prevenção de doenças, acarretando o aumento na demanda por produtos naturais, orgânicos e nutritivos (Esperancini, 2005).

3.1.2 Aspectos econômicos

Devido ao mercado altamente competitivo, as indústrias de bebidas estão sempre em busca da diversificação de sua linha de produtos, com destaque para os néctares de frutas, que representam a categoria do segmento de bebidas não alcoólicas com o maior aumento no consumo nos últimos anos (Neves *et al.*, 2012).

No cenário mundial, a categoria de bebidas à base de frutas, que engloba sucos integrais, néctares e refrescos, foi responsável por 4 % do mercado de bebidas não alcoólicas, do período de 2009 a 2013. Em 2013, o volume consumido foi de 23 bilhões de litros de sucos integrais, 14 bilhões de litros de néctares e 31 bilhões de litros de refresco. A maior parte do consumo foi da China (19 bilhões de litros) seguida dos Estados Unidos (12 bilhões de litros) (Ashurst, 2016).

Os maiores produtores de frutas são China, Brasil e Índia. Na China há uma produção significativa de maçã, no Brasil de laranja e na Índia de manga (Ashurst, 2016).

Dados da Associação Brasileira das Indústrias de Refrigerantes e de Bebidas Não Alcoólicas (ABIR) indicaram que, no ano de 2014, a produção brasileira de bebidas não alcólicas foi estimada em 36,6 milhões de litros, equivalentes a 180,3 litros per capita por ano. Dentre as bebidas listadas, se encontram refrigerante (44,7 %), água engarrafada (33,1 %), refresco em pó (13,7 %), néctares (3,6 %) sucos concentrados (1,8 %) e refrescos prontos para beber (1,4%) (ABIR, 2014). Foram consumidos 1,3 milhões de litros de néctar de fruta, um aumento de quase 15 % se comparado ao ano de 2013, correspondente a um consumo per capita de 6,4 litros por habitante por ano (ABIR, 2013).

Mundialmente, o néctar de laranja é o mais consumido (Neves *et al.*, 2012), enquanto no Brasil, são destaques de consumo os néctares de uva, pêssego,

laranja, maracujá, goiaba, caju e manga, nesta ordem (ABIR, 2015). **3.1.3**

Legislação

No Brasil, o registro, a padronização, a classificação e, ainda, a inspeção e a fiscalização da produção e do comércio de bebidas, em relação aos seus aspectos tecnológicos, competem ao Ministério da Agricultura, Pecuária e Abastecimento (MAPA) ou órgão estadual competente credenciado por esse Ministério, segundo a Lei nº 8.918 de 1994 (Brasil, 1994).

No Decreto nº 6.871, de 4 de junho de 2009, que regulamenta a Lei nº 8.918, de 14 de julho de 1994, estão descritos os tipos e as definições de bebidas não alcoólicas. Na Seção II, são definidos os tipos de suco: desidratado, misto, reconstituído e tropical, e também a designação do termo integral e a denominação concentrado para o suco parcialmente desidratado. No Art. 21º, é definido o termo néctar, como a bebida não fermentada, obtida da diluição em água potável da parte comestível do vegetal ou de seu extrato, adicionada de açúcares e destinada ao consumo direto (Brasil, 2009). De acordo com a legislação de aditivos alimentares, alguns aditivos, como acidulante ácido cítrico, antioxidante ácido ascórbico, estabilizante goma guar, são permitidos em néctares de frutas (ANVISA, 2013).

Em setembro de 2013, foi criada a Instrução Normativa (IN) nº42 que altera o Art. 3º da IN nº 12 de 2003 (MAPA, 2003), definindo quantidades mínimas de suco de 50 % (m/m) para os néctares de laranja e uva, a partir de 31 de janeiro de 2016 (MAPA, 2013). Além disso, com exceção do vegetal com acidez muito elevada ou sabor muito forte, o néctar cuja quantidade mínima da polpa de fruta ou do suco de fruta ou de vegetal não tenha sido fixada em Regulamento Técnico específico e o néctar misto devem conter, no mínimo, 30 % (m/m) da respectiva parte comestível do vegetal. Nestes casos, o conteúdo da polpa de fruta ou do suco de fruta ou de vegetal não deve ser inferior a 20 % (m/m). Nesta mesma IN é descrito que a quantidade da polpa de fruta ou do suco de fruta ou de vegetal, no néctar e no suco tropical, deve ser declarada no rótulo (MAPA, 2013).

Nas IN nº 12 de 2003 (MAPA, 2003) e IN nº 42 de 2013 (MAPA, 2013), são definidos os parâmetros do PIQ de diferentes néctares de frutas, incluindo o teor de suco ou polpa de fruta (g/100 g), sólidos solúveis (ºBrix), acidez total em ácido cítrico

(g/100 g), açúcares totais (g/100 g) e ácido ascórbico (mg/100 g). Na **Tabela 1** encontram-se relacionados os valores mínimos para os parâmetros do PIQ dos néctares das frutas que foram objeto de estudo desta tese.

Tabela 1 Padrões de Identidade e Qualidade (PIQ) dos néctares de caju, maracujá, pêssego, maçã, laranja e uva

Néctar	PIQ – mínimo				
	Suco ou polpa (g/100 g)	Sólidos solúveis (°Brix)	Acidez total em ácido cítrico (g/100 g)	Açúcares totais (g/100 g)	Ácido ascórbico (mg/100 g)
Caju	15	10	0,12	7	15
Maracujá	10	11	0,25	7	-
Pêssego	40	11	0,15	7	-
Maçã	30	-	-	-	-
Laranja	50	-	-	-	-
Uva	50	-	-	-	-

Fonte: MAPA (2003; 2013).

O Código de Defesa do Consumidor foi criado em 1990 (Lei nº 8.078) como forma de proteger a saúde do mesmo contra riscos advindos do consumo de produtos considerados perigosos ou nocivos. No Código é determinado que os produtos ou serviços ofertados devem assegurar informações corretas, claras, precisas, ostensivas e em língua portuguesa sobre suas características, qualidades, quantidade, composição, preço, garantia, prazos de validade e origem, entre outros dados, assim como sobre os riscos que apresentam à saúde e à segurança dos consumidores (Brasil, 1990).

O direito do consumidor de receber informações confiáveis sobre os alimentos objetiva a prevenção de práticas fraudulentas e enganosas, a adulteração de

alimentos e outras práticas que podem induzir ao erro (European Commission, 2002), sendo tais ações consideradas crimes contra a saúde pública (Brasil, 1940).

3.1.4 Métodos para detecção de fraudes em néctares

Bebidas à base de frutas são misturas complexas de substâncias como açúcares, ácidos orgânicos, compostos voláteis, aromas, ácidos graxos, aminoácidos, flavonoides e pigmentos em água. Assim, no contexto da adulteração em sucos e néctares, a sua detecção exige uma definição clara da composição do suco ou néctar, sendo que a autenticação destes produtos é de suma importância na indústria de alimentos, na qual a substituição com ingredientes mais baratos é uma prática comum que representa potencial ganho financeiro para as indústrias (Soria *et al.*, 2008; Asadpoor, Ansarin & Nemat, 2014).

Existem vários tipos conhecidos de adulteração em sucos e néctares que podem ocorrer durante o processo de fabricação, sendo que, quanto mais diluído for o produto, maiores as probabilidades de fraude. Os tipos de adulterações mais frequentes incluem: i) a simples diluição com água e a adição de xarope de açúcar; ii) o uso de corantes e ácidos orgânicos, que não estão naturalmente presentes nestes produtos; iii) a falsa declaração da origem geográfica do produto; iv) o uso de tecnologias não permitidas pela legislação (como no caso de cítricos, a adição de casca e/ou da polpa lavada); e v) a adição não declarada de diferentes espécies que podem ser ou não relacionadas botanicamente ao néctar de fruta em questão e/ou de perfil sensorial que não cause grande diferença nas características do produto final (Singhal, Kulkarni & Rege, 1997; Asadpoor, Ansarin & Nemat, 2014; Rinke, 2016).

Devido à composição química das bebidas à base de frutas e à ampla variação natural das mesmas, a adição de outras frutas, de menor custo ou com características sensoriais que não impactem no produto final, vem ocorrendo. Atualmente, é muito comum encontrar no mercado néctares mistos (que contém mais de uma fruta). Para garantir a autenticidade dos néctares que são vendidos

como “puros” (contendo apenas uma fruta) e evitar possíveis fraudes, a presença declarada de cada fruta deve ser confirmada no produto (Soria *et al.*, 2008).

Nos últimos dez anos, diferentes técnicas analíticas vêm sendo aplicadas à análise de fraudes em bebidas à base de frutas, conforme relacionado na **Tabela 2**.

Tabela 2 Técnicas analíticas para análise de fraudes em bebidas à base de frutas

Técnica	Tipo de análise	País	Referência
HPLC-DAD	Adulteração com limão, toranja e tangerina em suco de laranja	Espanha	Abad-García <i>et al.</i> , 2014
GC-MS	Adulteração com uva e pêssego em suco de romã	Espanha	Nuncio-Jáuregui <i>et al.</i> , 2014
HPLC-IRMS	Autenticação de sucos de limão e lima	França	Guyon <i>et al.</i> , 2014
PCR convencional	Detecção de uva, morango e laranja em suco de frutas	China	Han <i>et al.</i> , 2012
PCR em tempo real	Adulteração com tangerina em suco de laranja	Espanha	Pardo, 2015
Espectroscopia UV-VIS	Detecção de adição de água, suco de uva e suco de maçã em suco de romã	Itália	Boggia <i>et al.</i> , 2013
Espectroscopia FT-IR	Adulterações por diluição com xaropes de glicose, frutose e sacarose em sucos de laranja	Reino Unido	Ellis <i>et al.</i> , 2016
Espectroscopia FT-IR	Detecção de uva concórdia em <i>blends</i> de suco de uva	Estados Unidos	Snyder <i>et al.</i> , 2014

Legenda: FT-IR: infravermelho com transformada de Fourier; GC-MS: cromatografia gasosa acoplada à espectrometria de massas; HPLC-DAD: cromatografia líquida de alta eficiência com detector por arranjo de diodos; HPLC-IRMS: cromatografia líquida de alta eficiência acoplada à espectrometria de massa de razões isotópicas; PCR: reação em cadeia da polimerase; UV-VIS: ultravioleta-visível.

Algumas das técnicas descritas, como HPLC, GC-MS e PCR, são morosas, relativamente caras, com alto consumo de solventes e geração de resíduos. A vantagem do uso de técnicas espectroscópicas é a possibilidade de avaliar a composição dos alimentos empregando pequenas quantidades de amostra, utilizando métodos verdes, não destrutivos, simples, rápidos e diretos (Fernández-González *et al.*, 2014; Kirtil *et al.*, 2017).

Dos trabalhos encontrados na literatura, citados anteriormente, todos foram para aplicação na matriz suco, restritos à Europa, China e Estados Unidos e aos sabores de laranja, uva e romã, não cobrindo a matriz néctar.

3.2 TÉCNICAS ESPECTROSCÓPICAS

As técnicas espectroscópicas mais utilizadas para obter impressões digitais de alimentos são as espectroscopias de RMN, infravermelho próximo, infravermelho com transformada de Fourier (FT-IR) e UV-VIS. Estas técnicas possuem algumas vantagens para a análise da estrutura e composição dos alimentos, sendo possível analisar pequenas quantidades da amostra ou do seu extrato de forma não destrutiva, simples, rápida e direta, sem ou com mínima preparação de amostra, além da determinação simultânea de um grande número de compostos (Zhang *et al.*, 2011; Esslinger, Riedl & Fauhl-Hassek, 2013; Fernández-González *et al.*, 2014; Kirtil *et al.*, 2017). O uso destas técnicas atende à tendência da análise de alimentos com métodos verdes que geram respostas rápidas sem consumo de reagentes ou solventes e com menor impacto ambiental.

Um produto alimentar exposto à radiação do IV apresentará um espectro característico que será, essencialmente, o resultado da absorção por vários componentes químicos, embora as propriedades físicas do alimento possam também ter um impacto sobre os valores de absorbância no IV. Por causa da variação presente em qualquer material, a composição exata de um determinado lote ou amostra irá variar um pouco. No caso dos néctares de frutas, a composição irá depender da variedade da fruta, da estação do ano e da localização. Por conseguinte, haverá um perfil típico de espectros para cada alimento. Embora simples em termos de conceito, esta comparação não é trivial, sendo necessárias técnicas quimiométricas para a sua realização (Manley, Downey & Baeten, 2008).

3.2.1 Espectroscopia ATR-FTIR

Na espectroscopia no IV, cada grupo funcional apresenta uma região característica de absorção de acordo com seu modo de vibração molecular. Assim, o espectro total pode ser considerado uma impressão digital para um determinado composto (Fernández-González *et al.*, 2014). A espectroscopia de absorção molecular no IV é uma técnica ideal para triagem rápida e caracterização da composição química. Seu potencial aumenta quando a referida técnica é combinada

à quimiometria, se tornando robusta para a avaliação da qualidade e autenticidade de alimentos (Downey, 1998; Edelman *et al.*, 2001; He, Rodriguez-Saona & Giusti, 2007).

Ao analisar uma amostra, um feixe de luz IV passa através desta e a energia absorvida em cada número de onda é registrada. Isto pode ser feito de duas formas. A primeira é explorando todo o espectro com um feixe monocromático, modificando os comprimentos de onda com o tempo. A segunda utiliza o sistema de transformada de Fourier que mede todos os comprimentos de onda ao mesmo tempo. Assim, uma impressão digital (espectro) da amostra é obtida, auxiliando na confirmação da identidade da amostra (Karoui, Pierna & Dufour, 2008).

O espectro no infravermelho médio, situado entre 4000 cm^{-1} e 400 cm^{-1} , é composto de bandas bem resolvidas, relativas às vibrações fundamentais das moléculas, as quais podem fornecer informações importantes sobre os grupos funcionais presentes, diminuindo as possibilidades de atribuições espectrais ambíguas (Wilson, 1990; Reid, O'Donnell & Downey, 2006).

O uso de acessórios de reflexão foi o grande responsável pela ampliação do uso de métodos baseados no IV, devido à versatilidade e simplicidade de obtenção dos espectros. A análise de amostras realizada sobre um cristal de ATR é aplicável a qualquer tipo de amostra (sólida, líquida, pós, pastas, fibras, etc.), não é destrutiva e não demanda diluição. Ainda, há pouca variação induzida pelo operador e a leitura é feita em poucos segundos (Wilson, 1990; Karoui, Pierna & Dufour, 2008).

O uso de acessório de ATR, juntamente com métodos quimiométricos, propiciou, então, o desenvolvimento de métodos quantitativos e de classificação sem a necessidade de separar física e quimicamente os interferentes (Karoui, Pierna & Dufour, 2008).

3.2.2 Espectroscopia RMN-DT

A espectroscopia de RMN é uma técnica que envolve a análise da absorção de energia por núcleos atômicos com *spins* diferentes de zero, na presença de um

campo magnético. As absorções de energia dos núcleos atômicos são afetadas pelos núcleos de moléculas vizinhas, que causam pequenas modificações locais no campo magnético externo. A espectroscopia de RMN pode, portanto, fornecer informações detalhadas sobre a estrutura molecular de uma amostra de alimento baseada nas interações observadas de um núcleo atômico individual, que são dependentes dos átomos que o rodeiam (Reid, O'Donnell & Downey, 2006).

A espectroscopia de RMN de alta resolução, que utiliza frequências acima de 100 MHz, é aplicada em mais estudos de autenticidade de alimentos do que a espectroscopia de RMN de baixa resolução, a qual utiliza frequências de 10 MHz a 40 MHz. A espectroscopia de RMN de alta resolução permite a obtenção de informações mais detalhadas sobre a estrutura molecular de uma amostra de alimento. Porém, trata-se de uma das técnicas analíticas mais caras, devido aos custos iniciais de capital (equipamento) e aos seus custos de manutenção (Reid, O'Donnell & Downey, 2006).

Avanços na espectroscopia de RMN tornaram possível a criação de espectrômetros de volume e custo reduzidos, quando comparados aos de alto campo. O equipamento é chamado de RMN-DT, porque os sinais são obtidos monitorando os tempos de relaxação longitudinal (T_1) e transversal (T_2) (Colnago & Andrade, 2016). Ele usa baixo campo magnético e ímãs permanentes, na forma de um equipamento de bancada que é acessível, rápido e fácil de usar (Kirtil *et al.*, 2017).

O tempo de relaxação transversal T_2 , também chamado de *spin-spin*, se relaciona à perda de coerência da magnetização no plano $x'y'$ que é atribuída às interações diretas entre os momentos magnéticos individuais dos spins em uma amostra. Um decaimento exponencial da magnetização transversal é provocado pela relaxação transversal (Colnago & Andrade, 2016).

O tempo de relaxação longitudinal T_1 , também chamado de *spin-rede*, se relaciona ao retorno da magnetização ao estado de equilíbrio térmico. Essa relaxação resulta da existência de momentos magnéticos transitórios causados por movimentos rotacionais e translacionais das moléculas vizinhas, oscilando com

frequências adequadas em torno da frequência de Larmor (Colnago & Andrade, 2016).

Os sinais de RMN neste espectrômetro não necessitam de transformada de Fourier. Como são utilizados imãs de baixo campo e de baixa homogeneidade, os espectros obtidos são de baixa resolução, não sendo observado efeito do deslocamento químico ou de qualquer outra interação que resolva o espectro em mais de uma linha. Isso significa que, independente da amostra, só será observada uma única linha e, portanto, nos resultados obtidos com RMN-DT analisa-se a amplitude do sinal do decaimento livre da indução (*free induction decay* – FID) ou a recuperação do referido sinal na forma de eco em função do tempo (Colnago & Andrade, 2016).

3.3 QUIMIOMETRIA

De acordo com a Sociedade Internacional de Quimiometria, quimiometria é a ciência que relaciona medições feitas em um sistema ou processo químico ao estado do sistema através da aplicação de métodos matemáticos ou estatísticos (Hibbert *et al.*, 2009).

A quimiometria pode ser dividida, com relação aos seus objetivos, em aplicações de planejamento e otimização de experimento, de calibração multivariada e de reconhecimento de padrões (métodos de análise exploratória, classificação e resolução de curvas) (Barros Neto, Scarminio & Bruns, 2006), podendo ser aplicada a diversas áreas.

Inicialmente, a maioria dos trabalhos da literatura utilizando quimiometria tratava dados obtidos em química analítica instrumental. A maioria das aplicações envolvia técnicas como espectroscopia de absorção no UV/VIS, espectroscopia no IV e cromatografia. A partir dos anos 90, a quimiometria foi se expandindo para outras áreas de aplicação, como biologia e medicina, juntamente com aplicações de outras técnicas analíticas, como espectrometria de massas, RMN, etc. O uso da quimiometria em análises de alimentos para detecção de adulterações e fraudes e também em aplicações forenses utilizando informações químicas e espectroscópicas para determinar, por exemplo, a origem de amostras, também vem crescendo significativamente. A interface da quimiometria com a bioinformática é mais recente, estando relacionada a dados biológicos, havendo, também, aplicações na área de metabolômica (Brereton, 2007).

Neste trabalho, foram aplicadas a calibração multivariada utilizando o método PLS, e a classificação supervisionada utilizando os métodos SIMCA, PLS-DA, PLS-DM, DD-SIMCA e OCPLS.

3.3.1 Calibração multivariada

Na construção de um modelo de calibração multivariada, é estabelecida uma relação entre dois blocos de dados. Dois conjuntos de dados são correlacionados.

Em modelos de calibração inversa, como regressão de componentes principais e PLS, o primeiro conjunto, representado pela matriz \mathbf{X} , contém as medidas instrumentais que são as variáveis independentes. O segundo conjunto contém as propriedades de interesse a serem previstas pelo modelo, por exemplo, a concentração de um analito. Esse conjunto pode ser representado por um vetor \mathbf{y} (se for uma única variável prevista), ou por uma matriz, \mathbf{Y} (se forem mais de uma variável), de variáveis dependentes (Ferreira *et al.*, 1999; Ferreira, 2015).

Dois etapas são realizadas no processo de calibração. A primeira, que é a etapa de modelagem, em que é estabelecida uma relação matemática entre \mathbf{X} e \mathbf{Y} para um grupo de amostras representativas contidas no chamado conjunto de calibração. A segunda etapa, que é a etapa de validação ou previsão, em que o modelo construído é utilizado para prever a concentração das novas amostras independentes (Ferreira *et al.*, 1999; Ferreira, 2015).

3.3.1.1 Regressão por Mínimos Quadrados Parciais (PLS)

O método PLS foi desenvolvido nos anos 1960 pelo estatístico e econométrico Herman Wold (Wold, Sjöström & Eriksson, 2001).

O PLS é um método de regressão multivariada baseado na decomposição dos dados em variáveis latentes (VLs). As variáveis independentes e dependentes são decompostas, simultaneamente, de maneira similar à decomposição de componentes principais, antes de se estabelecer uma regressão entre os escores obtidos dessa decomposição (Ferreira *et al.*, 1999; Poppi & Sena, 2000). As VLs são as combinações lineares das variáveis em \mathbf{X} que melhor modelam, simultaneamente, as variáveis dependentes em \mathbf{Y} (Wold, Sjöström & Eriksson, 2001). Na decomposição em componentes principais, os dados espectrais (\mathbf{X}) são irrestritamente ortogonais. Com a decomposição simultânea de \mathbf{X} e \mathbf{Y} , há uma pequena perda da ortogonalidade das componentes principais, que, desta forma, recebem a denominação mais abrangente de VLs (Ferreira *et al.*, 1999; Poppi & Sena, 2000).

Na construção de modelos PLS, \mathbf{X} e \mathbf{Y} são decompostas, simultaneamente, em uma soma de h VLs (que descrevem a direção de máxima variância e se correlacionam com a concentração). \mathbf{X} é uma matriz $n \times p$ - em que n é o número de amostras e p é o número de variáveis independentes medidas; \mathbf{Y} pode ser um vetor (\mathbf{y}) $n \times 1$, quando o modelo PLS prediz apenas um analito por vez (PLS1), ou uma matriz $n \times k$, quando o PLS prediz k analitos simultaneamente (PLS2). A denominação PLS1 ou PLS2 não está relacionada com o número de analitos preditos, mas sim com a ordem do tensor dos valores de referência, pois um vetor é um tensor de primeira ordem, enquanto uma matriz é um tensor de segunda ordem (Geladi & Kowalski, 1986; Ferreira *et al.*, 1999).

De acordo com Geladi & Kowalski (1986), a decomposição das matrizes \mathbf{X} e \mathbf{Y} segue as **Equações 1 e 2**:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E}$$

(Equação 1)

$$\mathbf{Y} = \mathbf{UQ}' + \mathbf{F}$$

(Equação 2)

em que \mathbf{T} e \mathbf{U} , e \mathbf{P} e \mathbf{Q} são matrizes de escores e pesos (*loadings*) de \mathbf{X} e \mathbf{Y} , respectivamente, e as matrizes \mathbf{E} e \mathbf{F} são os correspondentes resíduos de \mathbf{X} e \mathbf{Y} .

Na **Figura 1** está representada esquematicamente a decomposição do modelo PLS (Geladi & Kowalski, 1986).

Figura 1 Decomposição dos dados pelo método dos mínimos quadrados parciais (PLS)

$$\begin{array}{c}
 \begin{array}{c} p \\ \boxed{\mathbf{X}} \\ n \end{array} = \begin{array}{c} h \\ \boxed{\mathbf{T}} \\ n \end{array} \begin{array}{c} p \\ \boxed{\mathbf{P}} \\ h \end{array} + \begin{array}{c} p \\ \boxed{\mathbf{E}} \\ n \end{array} \\
 \\
 \begin{array}{c} m \\ \boxed{\mathbf{Y}} \\ n \end{array} = \begin{array}{c} h \\ \boxed{\mathbf{U}} \\ n \end{array} \begin{array}{c} m \\ \boxed{\mathbf{Q}} \\ h \end{array} + \begin{array}{c} m \\ \boxed{\mathbf{F}} \\ n \end{array}
 \end{array}$$

Fonte: Adaptado de Geladi & Kowalski (1986).

O vetor de coeficientes de regressão linear (\mathbf{b}) correlaciona os blocos \mathbf{X} e \mathbf{Y} de forma linear, para h VLs, como descrito na **Equação 3**.

$$\mathbf{u}_h = \mathbf{b}_h \mathbf{t}_h$$

(Equação 3)

em que \mathbf{u}_h e \mathbf{t}_h denotam as h -ésimas colunas de \mathbf{U} e \mathbf{T} .

Agrupam-se os valores de \mathbf{b}_h na matriz \mathbf{B} , para o caso de previsão de mais de um analito (PLS2), contendo os coeficientes de regressão entre as matrizes de escores \mathbf{T} e \mathbf{U} (vetor \mathbf{u} no caso de PLS1). Obtém-se a melhor relação linear entre os escores por meio de pequenas rotações das VLs dos blocos \mathbf{X} e \mathbf{Y} . Calculam-se os valores preditos para as novas amostras ($\hat{\mathbf{Y}}$) com base nos escores (\mathbf{T}^*), seguindo a **Equação 4** (Geladi & Kowalski, 1986):

$$\hat{Y} = T^*BQ$$

(Equação 4)

Ao se construir modelos de calibração, dividem-se as amostras em conjuntos de calibração e validação. O conjunto de calibração deve ser representativo de toda a variância que se deseja modelar. O conjunto de validação deverá ser composto de amostras homogeneamente distribuídas na faixa de trabalho, sem extrapolar a amplitude da faixa do conjunto de calibração (Ferreira *et al.*, 1999).

3.3.2 Classificação multivariada

Os métodos de reconhecimento de padrões podem ser aplicados para objetivos distintos, como por exemplo, a análise exploratória de dados, a classificação de amostras e a resolução de curvas (Barros Neto, Scarminio & Bruns, 2006).

Os métodos de classificação podem ser divididos em dois tipos. Métodos não supervisionados, que buscam reduzir a dimensionalidade dos dados sem qualquer treinamento pré-definido, ou seja, sem o fornecimento prévio de nenhuma informação de atribuição de categoria/classe das amostras no modelo. Já os métodos de classificação supervisionados buscam classificar objetos em classes pré-definidas de acordo com suas características. Para isso, as amostras devem ser divididas em um conjunto de treinamento representativo e um conjunto de teste formado por amostras independentes (Brereton, 2015).

Ao se analisar uma amostra, avaliam-se os dados químicos da mesma, que consistem de certo número de objetos, descritos por um determinado número de variáveis. Em análises espectroscópicas, os objetos são os espectros de cada amostra e as variáveis são as absorvâncias em diferentes comprimentos/números de onda. A busca pela detecção de padrões de associação no conjunto de dados é utilizada para se estabelecer relações entre objetos e variáveis, para identificar

outliers (amostras ou variáveis anômalas) ou para agrupar objetos, podendo ser feita por meio da análise exploratória dos dados. Os métodos de análise exploratória mais utilizados são a análise de componentes principais (PCA) e a análise de agrupamentos hierárquica (HCA) (Barros Neto, Scarminio & Bruns, 2006).

Na PCA, procura-se reduzir a dimensionalidade do conjunto de dados e, ao mesmo tempo, preservar o máximo da informação. Para tanto, são estimadas combinações lineares das variáveis originais. A primeira componente principal, PC1, é a combinação linear de máxima variância, que modela o máximo de informação. A segunda componente, PC2, é a que possui a segunda maior variância, e é ortogonal à PC1 (portanto, independente). A terceira componente possui a terceira maior variância e é ortogonal às duas primeiras PCs e, assim, sucessivamente. As informações mais relevantes ficam concentradas nas primeiras PCs, que são utilizadas para procurar e definir padrões (Poppi & Sena, 2000; Barros Neto, Scarminio & Bruns, 2006).

Dentro do reconhecimento de padrões está incluído o conceito de classificação, que significa atribuir uma determinada amostra a uma ou mais categorias/classes, baseando-se em medidas feitas para caracterizar as amostras. Então, uma classe/categoria engloba amostras que possuam características similares, sendo que a definição destas características é problema-dependente (varia de acordo com o escopo do modelamento). As categorias são, portanto, definidas *a priori* e com os dados medidos com as amostras de treinamento constroem-se modelos matemáticos e estatísticos para atribuir as amostras desconhecidas às classes/categorias previamente definidas (Bevilacqua *et al.*, 2013).

Os métodos multivariados de classificação supervisionada podem ser divididos em dois tipos: modelagem de classes (classe única e multiclases) e discriminantes.

Métodos puramente discriminantes definem delimitadores no espaço hiperespaial das variáveis, dividindo-os em um número de regiões correspondente ao número de classes, sendo que o foco está na diferença entre as amostras das classes estudadas. Assim, a principal desvantagem deste tipo de método é que

amostras desconhecidas sempre serão classificadas como pertencentes a uma das classes modeladas. Métodos de modelagem de classes consideram, individualmente, cada classe, não levando em conta informações relativas às outras classes, sendo que a essência do método está na similaridade entre as amostras da mesma classe (Marini, 2010).

Os métodos de classe única são adequados quando o objetivo é testar se uma amostra é adulterada, independentemente de qual adulterante possa estar presente (López *et al.*, 2014). Quando o adulterante é conhecido, a estratégia multiclases pode ser escolhida, uma vez que fornece informações adicionais sobre as amostras, como múltiplas atribuições e não atribuição a nenhuma classe (Gondim *et al.*, 2017). Nos últimos anos, alguns autores têm criticado a predominância na literatura quimiométrica do uso de métodos discriminantes, tais como PLS-DA, para problemas de autenticação de alimentos (Rodionova, Oliveri, & Pomerantsev, 2016; Oliveri, 2017). Essa crítica se deve ao fato de que os resultados de classificação não seriam confiáveis quando o modelo for usado para prever uma nova amostra de uma classe não treinada. Em resposta, outros autores combinaram PLS-DA com detecção de *outliers*, identificando amostras de classes não treinadas baseadas em grandes valores de T^2 de Hotelling e resíduos Q (Martins *et al.*, 2017). No entanto, como os modelos de classe única são desenvolvidos usando apenas informações relativas às amostras de uma classe de cada vez, eles são incapazes de assegurar a seletividade do modelo para a detecção de várias fraudes alimentares (Xu *et al.*, 2016). Considerando todas essas discussões relevantes, é importante comparar as alternativas para o desenvolvimento de modelos de classificação supervisionada para detecção de fraude em alimentos.

3.3.2.1 Modelagem independente e flexível por analogia de classe (SIMCA)

Um dos modelos de classificação mais importantes é o SIMCA. Este modelo é baseado em componentes principais, que são as direções de máxima variância dentro de um espaço de dados multivariados (Oliveri & Downey, 2012; Bevilacqua *et al.*, 2013). O SIMCA pode ser utilizado como método de modelagem de uma única classe, quando uma única classe de amostras autênticas é modelada, ou como

método de modelagem de várias classes, quando pelo menos duas classes significativas são definidas no conjunto de dados (Oliveri & Downey, 2012).

No método SIMCA um modelo PCA é construído para cada classe de forma independente. Os limites das classes, estimados como valores de T_{im}^2 de *Hotelling* e resíduos Q_{im} , são calculados em um nível de significância específico (α), geralmente estabelecido em 0,05 (95 %). Os valores de T^2 de *Hotelling* e resíduos Q são usados para medir a informação de cada amostra incluída ou não no modelo (Rius, Callao & Rius, 1997).

Historicamente, diferentes critérios têm sido utilizados para a classificação das amostras em modelos SIMCA. Um critério comum de atribuição da amostra é baseado em seus valores de estatísticas reduzidas de T_r^2 de *Hotelling*, parâmetro relacionado com a influência de cada amostra no modelo, e Q_r , parâmetro relacionado aos resíduos do modelo. Esses valores reduzidos são as razões entre os parâmetros originais e os limites estatísticos para cada classe. Para ser considerada como "dentro do modelo" uma amostra deve ter valores inferiores a 1,0 para ambas as estatísticas reduzidas. Outro critério para a atribuição da amostra é uma variação do primeiro. Uma amostra i é atribuída baseada na sua distância para a classe j (d_{ij}), que é definida como a combinação dos parâmetros reduzidos da **Equação 5**. Assim, o limite da classe para uma amostra a ser considerada como "dentro do modelo" é um semicírculo com um raio 1 (d igual ou menor que 1). Portanto, este critério é mais restritivo do que considerar as estatísticas de T_r^2 de *Hotelling* e resíduos Q_r de forma independente (Márquez *et al.*, 2016):

$$d_{ij} = \sqrt{(Q_{r,i})^2 + (T_{r,i}^2)^2}$$

(Equação 5)

No modelo SIMCA desenvolvido neste trabalho, cada classe foi modelada independentemente das outras. Os valores de T^2 de *Hotelling* e Q para cada

amostra do conjunto de treinamento da classe alvo foram extraídos e utilizados para o cálculo do parâmetro d . Para as amostras da classe alvo, o valor de d menor que 1 indica que esta pertence à classe, enquanto um valor maior que 1 indica que ela não pertence (resultado falso-negativo). Para as amostras que não pertencem à classe alvo, o valor de d menor que 1 indica que elas pertencem à classe (resultado falso-positivo), uma vez que o valor esperado deve ser maior que 1 (não pertencem a classe) (Márquez *et al.*, 2016).

3.3.2.2 Análise discriminante por mínimos quadrados parciais (PLS-DA)

O objetivo do PLS-DA é muito similar ao do método de classificação análise discriminante linear (*Linear Discriminant Analysis* - LDA). Barker & Rayens (2003) mostraram que o PLS-DA é essencialmente o inverso da abordagem de mínimos quadrados para LDA e produz, essencialmente, o mesmo resultado. Porém com as vantagens do PLS em relação à robustez, à presença de colinearidades nos dados e à presença de ruído (Wise *et al.*, 2006).

O PLS-DA é, portanto, um método discriminante que adapta a técnica de regressão multivariada PLS a uma tarefa de classificação supervisionada, estabelecendo uma regressão linear entre uma matriz de variáveis independentes (\mathbf{X}) e uma matriz de variáveis dependentes (\mathbf{Y} ou \mathbf{y}). \mathbf{y} contém uma variável binária/categórica que indica a classe a que cada amostra pertence, sendo que 1 indica “pertencente” e 0 “não pertencente” (Barker & Rayens, 2003).

O modelo PLS-DA prediz a classe de cada amostra atribuindo valores entre 0 e 1. Como os valores previstos não são exatamente 0,0 ou 1,0, a estatística Bayesiana é usada para calcular o valor do limite (*threshold*) acima do qual a amostra é considerada pertencente à classe (Bylesjö *et al.*, 2006). O limite Bayesiano considera que os valores previstos do modelo PLS-DA seguem a distribuição normal, selecionando o valor de \mathbf{y} em que o número de resultados errôneos (falso-negativo e falso-positivo) é mínimo (Pulido *et al.*, 2003). Assim, os valores previstos acima ou abaixo deste limite definirão se a amostra pertence ou não à classe.

3.3.2.3 Mínimos quadrados parciais por modelagem de densidade (PLS-DM)

O PLS-DM é um método de classe única que adapta a regressão PLS a uma tarefa de classificação. Sua particularidade é determinar, para cada amostra do conjunto de treinamento, o vetor de resposta \mathbf{y} como uma estimativa da densidade de probabilidade da amostra, calculada como a soma das distâncias euclidianas entre k amostras com a menor distância no espaço multivariado (Oliveri *et al.*, 2014; Oliveri, 2017).

O valor de resposta (y_i) - para cada amostra i do conjunto de treinamento da classe a ser modelada - é calculado como uma estimativa da densidade amostral (d_i) baseado nas distâncias Euclidianas da amostra i em relação a cada uma das amostras de treinamento. As distâncias são ordenadas e d_i , obtida como a soma das menores distâncias k , é estudada variando k . Este parâmetro influencia a suavização da função de densidade, que evolui de um formato mais agudo para um mais suave, à medida que k aumenta (Oliveri *et al.*, 2014).

Depois da modelagem PLS, os escores das primeiras VLs são selecionados e utilizados como entrada para a estimativa das probabilidades de densidade da classe com diferentes coeficientes de suavização (a). A conformidade de cada objeto com o modelo de classes é observada ao satisfazer os critérios f_α e Q_α . " f_α " é o valor crítico da distribuição de densidade de probabilidade, em um nível de confiança pré-selecionado ($1 - \alpha$). " Q_α " são os resíduos PLS usados para calcular o valor crítico da estatística Q , no mesmo nível de confiança (Oliveri *et al.*, 2014).

O algoritmo utilizado aplica todas as combinações possíveis usando os parâmetros: distâncias k , coeficiente de suavização a (para a definição do espaço de classe no domínio dos escores PLS), número de VLs e pré-processamento adequado para a matriz \mathbf{X} . Então, a melhor combinação é escolhida com o ajuste do número de VLs por meio de critérios de eficiência (média geométrica de sensibilidade e seletividade) e com a avaliação dos outros parâmetros (Oliveri *et al.*, 2014; Oliveri, 2017).

Para este modelo, a seletividade é calculada na presença da classe não-alvo, podendo haver mais de uma classe externa. Nesse caso, a seletividade obtida é

calculada para as classes alternativas como um todo. Se a estimativa da seletividade de cada classe alternativa específica for necessária, ela deve ser calculada separadamente para cada classe não-alvo (Rodionova, Oliveri, & Pomerantsev, 2016).

3.3.2.4 Mínimos quadrados parciais de uma classe (OCPLS)

O OCPLS é um método de modelagem de classe. Nele, um modelo de mínimos quadrados parciais é ajustado correlacionando as medidas instrumentais com um vetor composto de valores 1 (Xu *et al.*, 2013). Ao contrário do SIMCA, cujos componentes principais são estimados explicando somente a variância das variáveis independentes, os componentes (VLs) do OCPLS consideram, simultaneamente, as variâncias explicadas nas variáveis independentes e dependentes e a compactação das amostras pertencentes à mesma classe (Xu *et al.*, 2014).

O número de VLs é estimado por validação cruzada. Quando um modelo OCPLS é construído duas medidas de distância são calculadas, o valor de *Hotelling T²* baseado na distância dos escores (SD) e os resíduos absolutos centralizados (ACR), conforme as **Equações 6 e 7** (Xu *et al.*, 2014):

$$ACR = |1 - \hat{y}_j - \hat{\mu}_e|$$

(Equação 6)

em que \hat{y}_j é a resposta ajustada do objeto e $\hat{\mu}_e$ é a média dos erros de treinamento.

$$T^2 = \sum_{i=1}^K \frac{(t_i - \bar{t}_i)^2}{s_{t,i}^2}$$

(Equação 7)

em que \bar{t}_i e $s_{t,i}^2$ são a média e variância da amostra da i -ésima VL, respectivamente, e, K é o número de VLs significativas.

O ACR pode assumir uma distribuição normal com média 0. O desvio padrão do modelo residual pode ser estimado por validação cruzada, de acordo com a **Equação 8** (Xu *et al.*, 2014):

$$\hat{\sigma}_e = \sqrt{\sum_{i=1}^N (1 - \hat{y}_i - \hat{\mu}_e)^2 / (N - 1)}$$

(Equação 8)

em que N é o número total de objetos omitidos durante a validação cruzada e \hat{y}_i é a resposta prevista do i -ésimo objeto omitido.

Dado um nível de confiança, α , os limites de confiança superiores (UCLs) para ACR e T^2 podem ser calculados, conforme as **Equações 9 e 10** (Xu *et al.*, 2014):

$$ACR_{UCL} = Z_{\alpha/2} \times \hat{\sigma}_e$$

(Equação 9)

$$T_{UCL}^2 = \frac{(n^2 - 1)K}{n(n - K)} F_{\alpha(K, n-K)}$$

(Equação 10)

em que $Z_{\alpha/2}$ é o ponto crítico superior da distribuição normal padrão e $F_{\alpha(K, n-K)}$ é o ponto crítico superior da distribuição F com $(K, n - K)$ graus de liberdade.

A SD é uma medida da distância de um objeto ao centro da classe no espaço ocupado pelas VLs significativas de OCPLS, e o ACR é uma medida de dispersão (dos resíduos) da projeção no vetor dos coeficientes de regressão OCPLS. De acordo com os valores de SD e ACR, são quatro as possíveis atribuições: amostra regular (pequeno SD e pequeno ACR); pontos de alavancagem (*leverage*) bons (grande SD e pequeno ACR); *outliers* de classe (pequeno SD e grande ACR); e pontos de alavancagem ruins (grande SD e grande ACR). Pontos de alavancagem bons, *outliers* de classe e pontos de alavancagem ruins podem ser detectados como três tipos diferentes de *outliers*, dependendo da definição dos problemas. Os limites para os valores de SD e ACR são estimados para níveis de confiança pré-estabelecidos (Xu *et al.*, 2014).

3.3.2.5 SIMCA orientado aos dados (DD-SIMCA)

O SIMCA orientado aos dados (*Data driven SIMCA*) é uma modificação do SIMCA original, a qual está principalmente relacionada à maneira de construir uma área de aceitação confiável para cada classe. Para este propósito, uma análise estatística completa das distâncias dos escores e das distâncias ortogonais é necessária. Os valores de distância devem ser considerados como quaisquer dados adquiridos no experimento, e suas distribuições são estimadas por um método orientado aos dados. A distribuição qui-quadrado fornece a possibilidade de construir uma área de decisão de dois níveis, com os limiares extremos e *outliers*, tanto no caso de um conjunto de dados regular quanto num conjunto contendo *outliers* (Pomerantsev & Rodionova, 2014a, b; Zontov *et al.*, 2017).

A primeira etapa da implementação do DD-SIMCA é a decomposição da matriz dos dados de treinamento utilizando a PCA, de acordo com a **Equação 11** (Zontov *et al.*, 2017):

$$X = TP^t + E$$

(Equação 11)

em que $T = \{t_{ia}\}$ é a matriz $(I \times A)$ de escores; $P = \{p_{ja}\}$ é a matriz $(J \times A)$ de pesos; $E = \{e_{ij}\}$ é a matriz $(I \times J)$ de resíduos e A é o número de PC.

Na segunda etapa, são calculadas a distância dos escores (SD), h_i , e a distância ortogonal (OD), v_i , para cada amostra do conjunto de treinamento com os resultados da PCA, conforme **Equações 12 e 13** (Zontov *et al.*, 2017):

$$h_i = t_i^t (T^t T)^{-1} t_i$$

(Equação 12)

$$v_i = \sum_{j=1}^J e_{ij}^2$$

(Equação 13)

A distância total para cada amostra é calculada conforme a **Equação 14**:

$$c = N_h \frac{h}{h_0} + N_v \frac{v}{v_0}$$

(Equação 14)

em que os parâmetros h_0 e v_0 são os fatores de escala, N_h e N_v são os números de graus de liberdade. Esses parâmetros são desconhecidos a priori e suas estimativas estão descritas nas referências Pomerantsev (2008) e Pomerantsev & Rodionova (2014a).

Na terceira etapa, é definida a área de aceitação para a classe alvo com base nos cálculos da distância dos escores e considerando os erros α e β (erros tipo I e II). O erro tipo I é a taxa de falsas rejeições, quando objetos da classe alvo são considerados não pertencentes a ela. O erro tipo II é a taxa de falsas aceitações, quando os objetos não pertencentes à classe alvo são classificados como pertencentes a ela (Pomerantsev & Rodionova, 2014b).

As amostras do conjunto de treinamento são caracterizadas por sua posição no gráfico de aceitação, podendo ser considerada uma amostra regular (atribuída à classe alvo) ou uma amostra extrema (localizada fora da área de aceitação e atribuída como classe não alvo) (Zontov *et al.*, 2017).

Além disso, é determinado um segundo nível de decisão para os *outliers*, construído para o valor γ fornecido. Tal valor especifica a probabilidade de que pelo menos um objeto regular do conjunto de dados seja considerado erroneamente como um *outlier*. Ao contrário da área de aceitação, a área de *outliers* depende do tamanho do conjunto de treinamento. Para um valor específico de γ , quanto maior for o tamanho deste conjunto, maior será a área de *outliers* (Zontov *et al.*, 2017).

Um gráfico de limiares extremos também é criado, de forma a demonstrar a dependência do número observado das amostras extremas *versus* os valores teoricamente esperados. Com o gráfico, é possível analisar a qualidade do modelo de classificação para o número de PC escolhido (Zontov *et al.*, 2017).

A última etapa é a classificação do conjunto de teste empregando o modelo definido. Os resultados da classificação são apresentados no gráfico de aceitação (Zontov *et al.*, 2017).

3.3.2.6 Algoritmo de Kennard-Stone

O algoritmo de Kennard-Stone foi desenvolvido no final dos anos 1960 e vem sendo muito utilizado como ferramenta para uma seleção adequada das amostras

para o conjunto treinamento/calibração na construção de modelos de regressão multivariados. O uso deste algoritmo visa à seleção das amostras de forma a garantir homogeneidade e representatividade, incluindo toda a variabilidade do conjunto de dados (Kennard & Stone, 1969).

A seleção das amostras se baseia nas distâncias Euclidianas no espaço multivariado, sendo que as duas primeiras amostras selecionadas para o conjunto de treinamento/calibração são aquelas que apresentam a maior e a menor distância em relação ao ponto médio central. A terceira amostra escolhida é a que apresenta a maior distância em relação às duas primeiras e, assim, o processo segue até que determinado número pré-estabelecido das amostras, normalmente 2/3 do total, seja selecionado (Kennard & Stone, 1969).

3.3.2.7 Validação cruzada

Uma etapa muito importante no processo de calibração multivariada é a definição do número de VLS. Dependendo desta escolha, pode haver casos de subajuste ou sobreajuste do modelo. No caso de subajuste, o número de VLS escolhido é insuficiente para descrever toda a variância sistemática a ser modelada, o que acaba por excluir informações relevantes do modelo. O sobreajuste (escolha de um número de VLS maior do que o adequado) causa problemas, porque não só incorpora características preditivas dos dados ao modelo, como também ruído. A implicação é um pior desempenho do modelo na fase de predição. Na prática, o sobreajuste é um dos problemas mais comuns na construção de modelos de calibração multivariada (Faber & Rajko, 2007).

Assim, como forma de evitar tanto sub como o sobreajuste, emprega-se a técnica de escolha de variáveis latentes chamada validação cruzada. A validação cruzada é considerada um processo de reamostragem, que determina o número de VLS com o qual o modelo apresenta menor RMSECV (raiz quadrada dos erros médios quadráticos de validação cruzada – em inglês *root mean square error of cross-validation*) (**Equação 15**) (Wise *et al.*, 2006; Faber & Rajko, 2007):

$$RMSECV = \sqrt{\frac{\sum(\hat{y}_i - y_i)^2}{n}}$$

(Equação 15)

em que \hat{y}_i corresponde aos valores de previsão das amostras que não foram incluídas na formulação do modelo e n ao número dessas amostras.

Em métodos de classificação multivariada, o melhor número de VL escolhido é o que apresentar o menor erro de classificação de validação cruzada (CVCE).

Os tipos de validação cruzada utilizados variam de acordo com a natureza dos dados e com os critérios de reamostragem. Para um conjunto pequeno de dados (com até 20 amostras de calibração), indica-se o *leave-one-out*, por meio do qual se retira uma amostra por vez. Para conjuntos de dados maiores, a validação por blocos contínuos é recomendada. Existem também as opções de subconjuntos aleatórios, para amostras organizadas de forma sistemática na matriz \mathbf{X} , e venezianas (*venetian blinds*), em que as amostras são retiradas em intervalos constantes (recomendado para conjuntos de amostras organizados em ordem sistemática de agrupamentos ou classes) (Ferreira, 2015).

3.3.4 Seleção de variáveis

A quantidade de informação gerada em análises espectroscópicas geralmente compreende centenas ou milhares de variáveis, e muitas delas podem ser ruidosas e/ou irrelevantes para o problema em estudo. A seleção de um número limitado de preditores informativos/variáveis pode melhorar a eficácia dos modelos de calibração e classificação, reduzindo sua complexidade e aumentando sua robustez (Andersen & Bro, 2010).

Tanto em modelos de calibração multivariada como em métodos de classificação, estratégias de seleção de variáveis podem ser empregadas para otimização (Andersen & Bro, 2010; Li *et al.*, 2016; Manfredi *et al.*, 2018). Existem métodos de seleção de variáveis que irão remover informações ruidosas e irrelevantes e outros que tem o objetivo de selecionar as variáveis mais discriminatórias ao trabalhar com diferentes grupos de amostras. Devido à ampla diversidade de métodos de seleção de variáveis, a escolha do mais apropriado não é uma tarefa simples (Di Anibal, Callao & Ruisánchez, 2011).

3.3.4.1 Regressão de mínimos quadrados parciais por intervalo (iPLS)

O iPLS é um dos métodos de seleção de variáveis mais comumente utilizados, uma vez que ele fornece uma visão geral do conjunto de dados, buscando descartar partes com informação irrelevante, interferências e ruídos. O espectro é dividido em um número de intervalos de comprimentos iguais e os modelos PLS são construídos para cada intervalo. Janelas de variáveis são indicadas quando os dados são altamente correlacionados, como dados espectrais (Andersen & Bro, 2010). A principal característica do iPLS é fornecer uma visão geral das informações relevantes em diferentes subdivisões espectrais, selecionando as regiões espectrais mais preditivas e removendo regiões com menor contribuição ao modelo. A sensibilidade do algoritmo PLS a variáveis ruidosas é destacada nos gráficos informativos obtidos pelo iPLS (Norgaard *et al.*, 2000).

A comparação visando encontrar o melhor modelo é baseada no menor valor de RMSECV, no número de VLS e na correlação dos valores medidos e preditos. O número de componentes é automaticamente escolhido e se baseia no menor valor de RMSECV (Norgaard *et al.*, 2000; Andersen & Bro, 2010).

3.3.4.2 Seleção dos preditores ordenados (OPS)

O OPS foi desenvolvido por Teófilo, Martins & Ferreira (2009). O objetivo desta metodologia é selecionar variáveis a partir de vetores informativos, seguindo

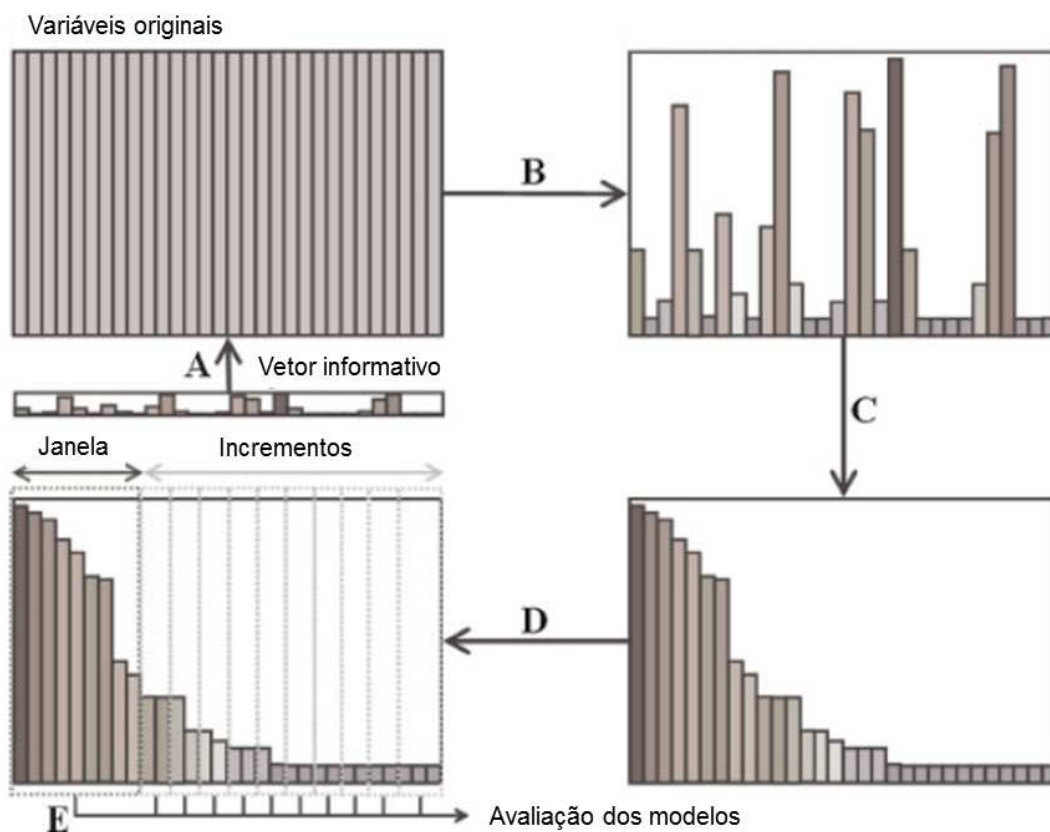
uma investigação sistemática de modelos de regressão PLS, e comparando os parâmetros de validação cruzada dos modelos obtidos a fim de encontrar o conjunto com as variáveis mais preditivas.

O método OPS pode ser descrito em quatro etapas: i) escolha do vetor informativo ou da combinação de vetores; ii) construção de modelos de regressão PLS; iii) cálculo dos parâmetros de qualidade utilizando validação cruzada; e iv) comparação dos parâmetros de qualidade obtidos em cada modelo.

O algoritmo utilizado no OPS é baseado em uma decomposição bidiagonal, e existem sete opções de vetores informativos para seleção, sendo estes: i) coeficientes de regressão; ii) vetor das correlações entre a matriz \mathbf{X} e o vetor \mathbf{y} ; iii) vetor de resíduos de predição; iv) vetor de covariância; v) vetor de VIP escores; vi) sinal analítico líquido (*net analyte signal* - NAS); e vii) vetor da razão sinal ruído.

Na **Figura 2** o método de seleção de variáveis OPS encontra-se representado esquematicamente. Inicialmente, o modelo de calibração é construído com os espectros originais inteiros (**Figura 2A**). Em seguida, são obtidos os vetores informativos a partir dos dados originais e as variáveis originais são sistematicamente diferenciadas de acordo com os valores fornecidos pelo vetor informativo (**Figura 2B**). As variáveis são organizadas em ordem decrescente (**Figura 2C**), de acordo com os valores absolutos dos elementos do vetor informativo, sendo que quanto maior o valor, maior a importância da variável. O próximo passo (**Figura 2D**) é a construção dos diferentes modelos de calibração multivariada utilizando validação cruzada do tipo *leave-N-out* (um grupo de N amostras é deixado de fora a cada vez). Seleciona-se uma janela (conjunto de variáveis) para construir o primeiro modelo otimizado e, assim, a matriz vai sendo expandida pela adição de um incremento (um certo número de variáveis). Desta forma, novos modelos são construídos e avaliados. Adicionam-se os incrementos até que uma parte ou todas as variáveis sejam avaliadas. Ao final, será selecionado o melhor modelo (com maior capacidade preditiva) (**Figura 2E**). Este será o que apresentar melhores parâmetros de qualidade da validação cruzada, menor valor de RMSECV e maior coeficiente de correlação (r) entre os valores previstos e de referência.

Figura 2 Representação esquemática do método seleção dos preditores ordenados (OPS)



Fonte: Adaptado de Teófilo, Martins & Ferreira (2009).

3.3.4.3 Algoritmo genético (GA)

Nos anos 60, o pesquisador da Universidade de Michigan, John H. Holland, propôs a construção de um algoritmo matemático para otimização em sistemas complexos, o GA. O referido algoritmo é um método de inteligência artificial utilizado em processos de otimização que se baseia na teoria da evolução das espécies, simulando, matematicamente, a teoria de Darwin (Costa Filho & Poppi, 1999).

O GA foi aplicado em quimiometria pela primeira vez por Lucasius & Kateman (1991) em um trabalho exploratório para seleção de comprimentos de onda em espectros no UV visando a análise de nucleotídeos.

A estrutura do GA permite que as informações referentes a um sistema de dados sejam codificadas de maneira análoga aos cromossomos biológicos, se assemelhando muito ao processo evolutivo natural. O GA envolve uma sequência de cinco etapas: i) codificação das variáveis; ii) criação da população inicial; iii) avaliação da resposta, iv) cruzamento e v) mutação (Costa Filho & Poppi, 1999; Leardi, 2001).

Na primeira etapa do método, é feita uma codificação em números binários (0 e 1) das variáveis (de maneira similar aos cromossomos), sendo que 0 significa que a variável será descartada e 1 que ela será incluída no modelo. Os melhores cromossomos que levarem aos melhores valores das figuras de mérito do modelo irão sobreviver, sofrer mutação e se recombinar para produzir descendentes. Após certo número de gerações, os melhores cromossomos serão selecionados e empregados para construir e prever o modelo (Goicoechea & Olivieri, 2003).

A criação da população inicial é feita de modo aleatório, formando a população inicial de cromossomos (combinação de todos os indivíduos testados) (Costa Filho & Poppi, 1999; Leardi, 2001).

Na etapa de avaliação, a mais importante no processo de GA, considera-se que a resposta seja a habilidade que um indivíduo possua para sobreviver, ou, em termos matemáticos, a habilidade de produzir a maior capacidade de previsão (com menor erro possível). Avalia-se esta resposta em relação ao erro, que é calculado pela diferença entre o valor esperado e o valor previsto. Os indivíduos que apresentarem menores erros serão os escolhidos para cruzar (Costa Filho & Poppi, 1999; Leardi, 2001).

Na quarta etapa, é realizado o cruzamento entre o material genético dos indivíduos. Na permuta, haverá maior tendência de transmissão das características dominantes para as novas gerações, assim, após algumas gerações, observa-se uma elevada taxa de indivíduos possuindo as variáveis preditivas. O cruzamento é responsável pela convergência para a situação desejada ótima (Costa Filho & Poppi, 1999; Leardi, 2001).

Na última etapa, ocorre uma perturbação nas informações contidas nos genes, causando a mutação. Estas alterações são causadas em uma pequena parte da população e, ao gerar indivíduos mais adaptados, as características serão transmitidas para as próximas gerações. Com isso, o problema de confinamento a soluções de mínimos locais é solucionado, pois esta etapa altera o direcionamento da busca pelo menor erro para outros locais da superfície de resposta (Costa Filho & Poppi, 1999; Leardi, 2001).

Um bom desempenho do GA depende da otimização de fatores como a probabilidade de mutação, a probabilidade de cruzamento, o tamanho da população inicial e o número de gerações. Para a probabilidade de mutação utiliza-se normalmente 1 %, enquanto para o cruzamento um valor em torno de 90 %. O tamanho da população inicial pode variar de 50 a 500 indivíduos e o número de gerações irá depender do tamanho da população inicial (Costa Filho & Poppi, 1999; Leardi, 2001).

3.3.4.4 VIP escores

Para a interpretação espectral de modelos PLS e PLS-DA normalmente são analisados os coeficientes de regressão do modelo. No entanto, a interpretação dos modelos não pode se basear somente nesses coeficientes, uma vez que eles são dependentes da composição das amostras no conjunto de treinamento/calibração, da covariância implícita entre os componentes dessas amostras e da relação sinal/ruído dos dados analíticos (Brown & Green, 2009).

Para avaliar de forma robusta e confiável, as variáveis importantes na projeção dos escores (VIP escores – *variable importance in projection*) fornecem uma ferramenta para interpretação espectral muito utilizada. Nos modelos PLS ou PLS-DA, a importância de cada variável na projeção é estimada através dos coeficientes de cada componente e pela significância de cada componente na regressão (valor em módulo). Assim, os VIP escores são considerados mais robustos que os coeficientes de regressão ao identificar quais variáveis são mais significativas para o modelo de predição, embora eles não indiquem a direção da

influência de cada variável (Wold, Sjöström & Eriksson, 2001; Chong & Jun, 2005; Brown & Green, 2009).

Ao serem aplicados para seleção de variáveis, recomenda-se estabelecer um limite (*threshold*) de 1,0, ou seja, selecionar apenas as variáveis que apresentem valores de VIP escores maior que 1,0 (Chong & Jun, 2005).

3.3.4.5 *iPLS* combinado a outro método de seleção de variáveis

O uso de método de seleção de variáveis discretas, tais como GA, OPS e VIP escores, combinado com a prévia aplicação de um método de seleção de variáveis contínuas, como o *iPLS* modificado, foi descrito na literatura (Leardi & Norgaard, 2004). A combinação de duas metodologias de seleção de variáveis pode melhorar a detecção e remoção de variáveis não relevantes ao modelo, reduzindo o domínio de pesquisa para um número de variáveis menor, o que pode melhorar o desempenho do modelo.

3.4 VALIDAÇÃO ANALÍTICA MULTIVARIADA

3.4.1 Abordagem quantitativa

Em experimentos de calibração é recomendado que a concentração de interesse esteja próxima ao centro da faixa de trabalho. Em protocolos de validação de dados univariados, na avaliação do desvio da linearidade, alguns autores sugerem o uso de cinco ou seis níveis de concentração sendo um nível central e os outros homoganeamente distribuídos até os extremos. O número de replicatas mínimo varia de dois a três. Recomenda-se que os níveis de concentração sejam equidistantes ao longo da faixa de estudo e, para garantir independência dos resultados, aconselha-se a leitura/medida das amostras em ordem aleatória (Thompson, Ellison & Wood, 2002; Souza & Junqueira, 2005; Souza, 2007).

Para avaliar a adequação do modelo linear (linearidade), após a detecção de *outliers*, as premissas de normalidade, homoscedasticidade e independência dos resíduos da regressão devem ser testadas. Pode-se avaliar a normalidade pelo teste de Ryan & Joiner, a homoscedasticidade pelo teste de Brown & Forsythe e a independência pelo teste de Durbin & Watson (Souza & Junqueira, 2005; Souza, 2007).

A avaliação da veracidade (a concordância entre os valores medidos e os valores de referência) pode ser feita com o uso de materiais de referência certificados, testes de recuperação ou métodos de referência. Na verificação da precisão (que avalia a proximidade de concordância entre resultados obtidos, sob condições previamente definidas, independente do valor verdadeiro ou especificado), os ensaios são replicados com amostras idênticas ou homogêneas, sob condições de repetibilidade e precisão intermediária, podendo estas serem expressas como desvios padrão relativos (European Commission, 2002; Thompson, Ellison & Wood, 2002; Souza, Pinto & Junqueira, 2007; Souza, 2007).

Alguns protocolos de validação quantitativa para dados multivariados citam também a estimativa de parâmetros como seletividade, sensibilidade (SEN),

sensibilidade analítica (γ), inverso da sensibilidade analítica (γ^{-1}), limite de detecção (LOD), limite de quantificação (LQ) e viés (*bias*) (Valderrama, Braga & Poppi, 2009; Botelho, Mendes & Sena, 2013).

A seletividade, sensibilidade, LD e LQ dependem da estimativa do sinal analítico líquido (NAS). O NAS pode ser definido como a parte do sinal analítico unicamente relacionado com o analito ou propriedade de interesse, o qual é ortogonal ao espaço dos interferentes (Botelho, Mendes & Sena, 2013). O vetor NAS ($\hat{x}_{A,i}^{nas}$) pode ser estimado para cada amostra (i) a partir do vetor de regressão (\mathbf{b}) do modelo PLS com h variáveis latentes, de acordo com a **Equação 16**:

$$\hat{x}_{A,i}^{nas} = \mathbf{b}(\mathbf{b}^T \mathbf{b})^{-1} \mathbf{b}^T x_i$$

(Equação 16)

A sensibilidade (SEN) em métodos multivariados é estimada como o valor de NAS para a concentração unitária (Valderrama, Braga & Poppi, 2009; Botelho, Mendes & Sena, 2013), de acordo com a **Equação 17**:

$$SEN = \frac{1}{\|\mathbf{b}\|}$$

(Equação 17)

em que “ $\|\ \|$ ” indica a norma Euclideana do vetor.

A γ é a razão entre a SEN e o ruído instrumental (ϵ) (**Equação 18**). O ruído instrumental é estimado pelo desvio padrão do vetor contendo 15 a 20 replicatas do espectro do branco (Valderrama, Braga & Poppi, 2009; Botelho, Mendes & Sena, 2013).

$$\gamma = SEN/\varepsilon$$

(Equação 18)

O γ^{-1} estima a diferença mínima discernível pelo método, considerando-se o erro aleatório instrumental como a única fonte significativa de erros, independente da técnica empregada (Valderrama, Braga & Poppi, 2009; Botelho, Mendes & Sena, 2013).

O viés tem relação com a presença de erros sistemáticos e é definido como a simples somatória de todos os erros de previsão, levando-se em conta os respectivos sinais (ASTM, 2012). Utiliza-se para o cálculo apenas as amostras de validação (Botelho, Mendes & Sena, 2013), de acordo com a **Equação 19**:

$$viés = \sum_{i=1}^{n_v} \frac{(y_i^{ref} - y_i)}{n_v}$$

(Equação 69)

em que y_i^{ref} é o valor de referência de concentração, y_i é o valor predito do modelo PLS e n_v é o número total de amostras no conjunto de validação. Para checar se o viés é significativo, realiza-se um teste t com 95 % de confiança, levando-se em conta o seu valor e o respectivo desvio padrão.

A razão de desempenho do desvio (RPD) é a razão da variação natural nas amostras de calibração ou validação em relação aos erros prováveis ocorridos durante a previsão (Valderrama, Braga & Poppi, 2009; Botelho, Mendes & Sena, 2013). As **Equações 20** e **21** são utilizadas para calcular os valores de RPD de calibração e de validação, respectivamente:

$$RPD_{cal} = DP_{cal}/RMSECV$$

(Equação 20)

$$RPD_{val} = DP_{val}/RMSEP$$

(Equação 71)

O desempenho do modelo pode se basear, também, nos valores de RPD, sendo que o um modelo de previsão pode ser considerado excelente quando o valor de RPD é maior do que 2,5. Previsões inferiores são obtidas quando o RPD está entre 2 e 2,5; e, entre 1,5 e 2, significa que o modelo pode discriminar valores baixos de altos da variável de resposta (Nicolai *et al.*, 2007).

3.4.2 Abordagem qualitativa

Parâmetros de desempenho (ou figuras de mérito - FOM) são atributos mensuráveis que indicam a qualidade do método analítico e, especificamente para modelos qualitativos, os parâmetros mais comumente utilizados são as taxas de sensibilidade (SEN), seletividade (SEL) e resultados inconclusivos (IR). Os dois primeiros são baseados em probabilidades com quatro possíveis respostas binárias: verdadeiro positivo (TP), verdadeiro negativo (TN), falso-positivo (FP) e falso-negativo (FN). Estas probabilidades são utilizadas para calcular os parâmetros de desempenho (López, Callao & Ruisánchez, 2015; Gondim *et al.*, 2017), de acordo com as **Equações 22, 23 e 24**:

$$SEN = TP_j/n^oS_j$$

(Equação 22)

sendo TP_j o número de amostras da classe j que foram corretamente preditas como pertencentes à classe j e n^oS_j o total de amostras que realmente pertence à classe j .

$$SEL = \frac{TN_j}{n^oS_{not j}}$$

(Equação 23)

sendo TN o número de amostras que não são da classe j que foram preditas como não pertencentes à classe j e $n^oS_{not j}$ o total de amostras que realmente não pertencem a classe j .

$$IR = (NA_j + MA)/n^oS_j$$

(Equação 24)

sendo NA_j o número de amostras que não foram atribuídas a nenhuma classe e MA o número de amostras da classe j que foram atribuídas a mais de uma classe.

No presente trabalho foi adotado o uso da expressão “seletividade” em detrimento de “especificidade”. O termo expressa a extensão na qual um método, em particular, pode ser utilizado para determinar analitos na presença de outros componentes de comportamento similar, sob determinadas condições (Vessman *et al.*, 2001).

A eficiência de um modelo qualitativo pode ser calculada como a média geométrica da sensibilidade e da seletividade, de acordo com a **Equação 25**:

$$EFF = \sqrt{SEN \times SEL}$$

(Equação 25)

4 MATERIAL E MÉTODOS

Os néctares foram preparados no Laboratório de Bromatologia - Unidade de Pesquisa e Análise de Alimentos (BRO-UPAA) e no Laboratório de Operações e Tecnologia (SOPT), ambos do Departamento de Alimentos (ALM) da Faculdade de Farmácia (FAFAR) da UFMG.

As análises por ATR-FTIR dos néctares foram realizadas no Laboratório de Biocombustíveis da Escola de Engenharia da UFMG e no Departamento de Química (DQ) da UFMG.

As análises por RMN-DT foram realizadas no Laboratório Nacional Agropecuário (LANAGRO-MG), em Pedro Leopoldo, MG.

4.1 EQUIPAMENTOS

Os equipamentos utilizados para processamento dos néctares incluíram: agitador magnético (Fisatom 752), aparelho de oxirredução por titulação (REDUTEC TECNAL TE 086), autoclave Vertical (FANEM Modelo 415), balanças analíticas calibradas (Gehaka BK300 e Shimadzu), liquidificador Poli (Metalúrgica Siemens LTDA, modelo LS-04), multiprocessador All in One BR2 (Philco), refratômetro (Hanna HI 96801) de 0 °Brix a 85 °Brix, refrigerador (Electrolux), pHmetro calibrado (Bante Instruments 920), termômetro digital calibrado (Incoterm), ultrapurificador de água (Millipore Direct-Q® 3 UV).

Para análise por ATR-FTIR foram utilizados o espectrofotômetro ATR-FTIR Frontier (Perkin Elmer) com acessório ATR de diamante – Espectrofotômetro 1 (IV 1) e o espectrofotômetro IRAffinity-1 FTIR (Shimadzu) com detector DLATGS (Deuterated Triglycine Sulfate Doped with L-Alanine) e acessório ATR horizontal contendo um prisma de ZnSe (PIKE Technologies, Madison, WI, EUA) de 20 reflexões internas – Espectrofotômetro 2 (IV 2).

Para análise por RMN foi utilizado o espectrômetro de RMN Specfit (Fine Instrument Technology, São Carlos, SP, Brasil) equipado com magneto nacional de 25 MHz (0,6 T).

4.2 MATERIAL

Os recipientes de acondicionamento das frutas, tábuas, facas e peneiras foram de material inerte e de primeiro uso. Para as análises espectroscópicas foram utilizados: algodão, pipeta de Pasteur, microtubo tipo Eppendorf (1,5 e 2,0 mL) e tubo Falcon (50 mL). Os demais materiais são descritos, a seguir, por análise.

4.2.1 Determinação de acidez total titulável e pH

Para a determinação de acidez total titulável foram utilizados: balões volumétricos (100 e 1000 mL), bastão de vidro e de plástico, béqueres de plástico (100 e 1000 mL), béqueres de vidro (100 e 250 mL), béqueres de vidro forma alta (250 mL), bureta com escala centesimal (10 mL), cápsula de porcelana, erlenmeyers (150 mL), etiquetas para identificação das soluções, frascos de polietileno para armazenar soluções, papel toalha, pisseta com água deionizada, ponteiras descartáveis (1 a 10 mL), provetas (100 mL), soluções tampão de pH 4,0 e pH 7,0, suporte para bureta.

4.2.2 Determinação de açúcares totais

Para a determinação de açúcares totais foram utilizados: balão volumétrico (100, 500 e 1000 mL), bandeja para banho de gelo, bastão de vidro e de plástico, béquer de plástico (100, 250 e 500 mL), béquer de vidro (100, 250, 500 e 1000 mL), bolsas de gelo, bureta (25 mL), erlenmeyer (125 e 250 mL), frasco para armazenamento de soluções, funil de vidro e de plástico, papel de filtro qualitativo 12,5 cm, pinça de metal, pipeta de Pasteur, ponteira descartável (1 a 10 mL), proveta (50, 100, 500 e 1000 mL) e pisseta.

4.2.3 Determinação do teor de ácido ascórbico

Para a determinação do teor de ácido ascórbico foram utilizados: béquer (50 mL), pipeta volumétrica (10 mL), bureta (10 mL), erlenmeyer ou balão de fundo chato (250 mL), balão volumétrico (100 e 1000 mL) e barra magnética.

4.2.4 Análise por espectroscopia ATR-FTIR

Para as análises de espectroscopia ATR-FTIR foram utilizados: algodão, pipeta de Pasteur, microtubo tipo *eppendorf* (1,5 e 2,0 mL) e tubo Falcon (50 mL).

4.3 REAGENTES, PADRÕES E OUTROS

Os reagentes utilizados em todas as análises apresentaram grau de pureza apropriado. Água ultrapura foi obtida utilizando o sistema Millipore Direct-Q[®] 3 UV (Billerica, MA, Estados Unidos).

4.3.1 Determinação de acidez total titulável e pH

Os reagentes fenolftaleína, álcool etílico foram adquiridos da Dinâmica Química Contemporânea Ltda. (Diadema, SP, Brasil), hidróxido de sódio e biftalato de potássio da Synth (Diadema, SP, Brasil) e hidróxido de bário da Cinética Reagentes e Soluções (Londrina, PR, Brasil). Cloreto de potássio foi fornecido pela Synth (Diadema, SP, Brasil) e fita indicadora de pH da Macherey-Nagel (Düren, Alemanha).

4.3.2 Determinação de açúcares totais

Acetato de chumbo, azul de metileno e hidróxido de sódio em micropérola foram obtidos da Vetec Química Fina Ltda (Rio de Janeiro, RJ, Brasil), D-glicose da Isofar (Duque de Caxias, RJ, Brasil) e sulfato de cobre pentahidratado e tartarato

duplo de sódio e potássio da Synth (Diadema, SP, Brasil). Ácido clorídrico a 37% foram fornecidos pela CRQ Produtos Químicos Industriais (Diadema, SP, Brasil), carvão ativo pela Synth (Diadema, SP, Brasil) e papel de tornassol vermelho pela Química Especializada Erich Ltda (São Paulo, SP, Brasil).

4.3.3 Determinação do teor de ácido ascórbico

Ácido oxálico foi adquirido da Synth (Diadema, SP, Brasil), ácido L-ascórbico da Neon Comercial Ltda. (São Paulo, SP, Brasil) e 2,6-diclorofenolindofenol-sódio da Vetec Quimica Fina Ltda (Rio de Janeiro, RJ, Brasil).

4.3.4 Análise por espectroscopia ATR-FTIR

Acetona para análise (p.a.) foi adquirida da Quimex (São Paulo, SP, Brasil).

4.3.5 Processamento dos néctares

Todas as matérias-primas utilizadas para a produção de néctares foram de grau alimentício. Os aditivos utilizados foram: acidulante ácido cítrico, antioxidante ácido ascórbico e estabilizante/espessante goma guar adquiridos da Pryme Foods (Sorocaba, SP, Brasil). Açúcar foi comprado da Camil Alimentos S/A (São Paulo, SP, Brasil). Hipoclorito de sódio utilizado para limpeza das frutas foi obtido da Vetec Quimica Fina Ltda (Rio de Janeiro, RJ, Brasil).

4.4 AMOSTRAS

4.4.1 Frutas

As variedades de cada fruta foram as mesmas utilizadas pelas indústrias nos preparos dos néctares, segundo ABIR (ABIR, 2015). Amostras de pêsego (*Prunus persica*) importado, laranja (*Citrus sinensis*) pêra rio, maracujá (*Passiflora edulis*) amarelo, caju (*Anacardium occidentale*) vermelho e maçã (*Malus domestica*) Fuji e Gala foram adquiridas do CEASAMINAS (Centrais de Abastecimento de Minas Gerais S.A.), em Contagem, MG. Amostras de uva (*Vitis labrusca*) Isabel foram fornecidas pela Embrapa Uva e Vinho, de Petrolina, PE.

A seleção das frutas levou em conta a ausência de danos mecânicos e fitopatológicos, o grau de maturação e outras características físicas típicas de cada fruta como tamanho, cor e textura (Paltrinieri & Figuerola, 1998). As frutas foram acondicionadas, sob refrigeração (4 a 7°C), até o momento do preparo dos sucos ou polpas para formulação dos néctares (**Tabela 3**).

Tabela 3 Descrição das frutas utilizadas no preparo dos néctares com respectivas variedades, cor, grau de maturação condições de refrigeração, branqueamento e rendimento

Fruta	Variedade	Cor	Grau de maturação	Temperatura de refrigeração (°C)	Branqueamento	Rendimento (%)*
Caju	Vermelho	Vermelho	Pedúnculo colhido quando completamente desenvolvido: tamanho máximo, firme e a coloração característica do tipo (Oliveira, 2007)	5 (Oliveira, 2007)	Não é necessário	59
Laranja	Pêra rio	Verde	Mínimo de suco deve estar entre 35 e 45 %; teor de sólidos solúveis entre 9 e 10 °Brix e relação (Sólidos Solúveis Totais/Acidez Total Titulável) entre 8,5 e 10 (Pereira <i>et al.</i> , 2006)	3 a 9 (Chitarra & Chitarra, 2005)	Não é necessário	45
Maçã	Fuji e Gala	Vermelho	Peso específico, sólidos solúveis, teor de amido e a razão entre o teor de ácidos e açúcares (1:10 a 1:15 dependendo da cultivar) (Esperancini, 2005)	-1 a 4 (Chitarra & Chitarra, 2005)	3 minutos a 100 °C e posterior banho de gelo	Maçã Fuji: 41 Maçã Gala: 35
Maracujá	Amarelo	Verde-amarelada	Conteúdo de sólidos solúveis totais entre 13 e 18 °Brix) (Cançado Júnior, Estanislau & Paiva, 2000)	7 a 10 (Chitarra & Chitarra, 2005)	Não é necessário	25

Pêssego	Importado Chile e Argentina	Cor de fundo amarelo- clara e polpa amarela ou laranja	Firmeza varia entre 11 lb (mínimo) e 17 lb (máximo). Para as cultivares de ciclo médio ou tardio, a quantidade de sólidos solúveis varia de 12 a 17 °Brix. Nas precoces esta quantidade varia entre 9 e 10 °Brix (Medeiros, 2002)	-0,5 a 0 (Chitarra & Chitarra, 2005)	3 minutos a 100 °C e posterior banho de gelo	59
Uva	Isabel	A cor das bagas varia do verde ao roxo	A uva deve estar macia e deve apresentar teor de sólidos solúveis totais igual ou superior a 15 °Brix (Leão, 2004)	-1 a -0,5 (Chitarra & Chitarra, 2005)	Não é necessário	42

Legenda: *Os valores de rendimento de cada fruta foram estabelecidos em experimentos piloto.

4.4.2 Néctares comerciais

Foram adquiridos, no comércio de Belo Horizonte (MG), néctares de uva, pêssego, laranja, maracujá, caju e maçã, de sete diferentes marcas (A, B, C, D, E, F e G) de cada fruta, todas de abrangência nacional, acondicionadas em embalagem Tetrapak de 1 L.

Cumprir destacar que não havia todas as marcas para todos os sabores estudados, uma vez que nem todos os fabricantes produzem néctares de todas as frutas analisadas. Para uva e caju todas as marcas foram avaliadas. Para laranja, foram avaliadas as marcas A, B, C, D e G; para pêssego, as marcas A, B, D, E, F e G; para maracujá, as marcas A, B, D, E e G; e para maçã, as marcas A, D, F e G.

Os néctares comerciais foram adquiridos no período de março a abril de 2016. As datas de validade das amostras estão relacionadas na **Tabela 4**:

Tabela 4 Datas de validade dos néctares comerciais adquiridos

Sabor	Marca*						
	A	B	C	D	E	F	G
Caju	17/10/2016	11/10/2016	09/03/2017	11/12/2016	22/12/2016	07/01/2017	25/11/2016
Laranja	24/10/2016	02/08/2016	30/12/2016	12/02/2017	-	-	27/08/2016
Maçã	25/09/2016	-	-	17/01/2017	-	13/09/2016	01/02/2017
Maracujá	19/02/2017	19/11/2016	-	21/12/2016	18/01/2017	-	18/03/2017
Pêssego	27/01/2017	30/10/2016	-	02/12/2016	30/12/2016	05/12/2016	23/03/2017
Uva	12/02/2017	09/11/2016	09/01/2017	24/01/2017	19/03/2017	01/03/2017	10/03/2017

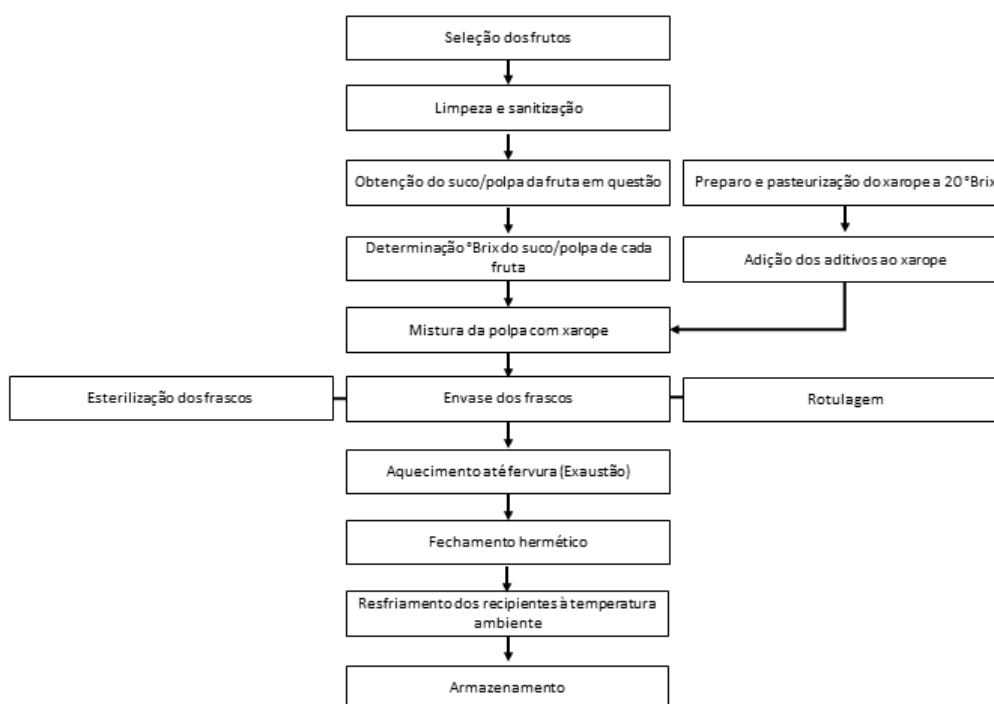
*Algumas marcas não produzem todos os sabores de néctares estudados.

4.4.3 Néctares formulados

Os néctares adulterados e não adulterados foram formulados em escala piloto, sendo otimizadas as condições de produção de cada néctar, em função do rendimento e do teor de açúcar ($^{\circ}$ Brix) dos sucos ou polpas definidos na **Tabela 3**.

Os néctares foram processados conforme fluxograma da **Figura 3** (adaptado de acordo com o rendimento da fruta em questão).

Figura 3 Fluxograma básico do processamento dos néctares



Fonte: Adaptado de Paltrinieri, Figuerola & Rojas (1997), Paltrinieri & Figuerola (1998) e Moraes (2006).

Inicialmente, os frutos foram retirados da refrigeração, deixando atingir temperatura ambiente para início do processamento. Foi feita uma inspeção visual

para seleção dos frutos dentro dos padrões definidos na **Tabela 3**. Os frutos foram limpos em solução de hipoclorito de sódio a 100 mg/L de cloro total por 2 min. Esta concentração foi padronizada levando em consideração a média encontrada na literatura para a sanitização das diferentes frutas (Nascimento Neto, 2006).

A etapa de obtenção do suco/polpa variou de acordo com a fruta em questão, conforme descrito a seguir:

- caju - a castanha foi retirada, o fruto foi cortado em quatro pedaços e despulpado em liquidificador industrial até ser transformado em polpa. Em seguida, a polpa obtida foi peneirada;
- laranja - a fruta foi cortada ao meio e as metades tiveram o suco extraído em espremedor. O suco foi peneirado para retirada dos gomos;
- maçã - a fruta foi descascada, cortada em oito pedaços e as sementes retiradas. Em seguida, ela foi branqueada em água fervente por 3 min seguido da imersão em água com gelo até resfriamento da fruta (máximo de 20 °C) (Paltrinieri, Figuerola & Rojas, 1997). A fruta branqueada foi despulpada em liquidificador industrial e peneirada;
- maracujá - a fruta foi cortada ao meio e as sementes peneiradas até retirada do suco aderido;
- pêsego - a fruta foi cortada em oito pedaços e o caroço removido. Em seguida, ela foi branqueada em água fervente por 3 min seguido da imersão em água com gelo até resfriamento da fruta (máximo de 20 °C) (Paltrinieri, Figuerola & Rojas, 1997). A fruta branqueada foi despulpada em liquidificador industrial e peneirada;
- uva - as uvas foram aquecidas em vapor constante por 1 h a 2 h (autoclave a 100 °C), prensadas e peneiradas para retirada do suco.

O despulpamento foi feito em liquidificador industrial (Metalúrgica Siemens LTDA) e o peneiramento em peneiras de 1 mm (Bates, Morris & Crandall, 2001). Os

graus Brix de cada suco/polpa produzido foram medidos utilizando refratômetro. O xarope foi preparado levando em conta o °Brix inicial do suco/polpa e o °Brix final do néctar que deveria variar entre 11 e 13 (conforme intervalo obtido de néctares comerciais). Foi utilizada a **Equação 26** para os cálculos:

$$\frac{a \times A}{100} + \frac{b \times B}{100} + \frac{c \times C}{100} + \frac{d \times D}{100} = \frac{m \times (A + B + C + D)}{100}$$

(Equação 26)

em que "a", "b", "c" e "d", correspondem aos valores de Brix da fruta principal, do xarope, da fruta adulterante 1 e da fruta adulterante 2, respectivamente; "A", "B", "C" e "D" representam a porcentagem de fruta principal, xarope, fruta adulterante 1 e fruta adulterante 2, respectivamente. "m" é o Brix final do néctar e "A + B + C + D" é igual a 100 (Tressler & Joslyn, 1961). Sendo que, se nenhum adulterante for adicionado, essa fórmula compreenderá apenas "A" e "B". Se apenas um adulterante for adicionado, a fórmula compreenderá "A", "B" e "C".

Ao xarope foram adicionados os aditivos em concentração que resulte em um produto final com parâmetros que cumpram com os requisitos mínimos do PIQ de cada néctar, incluindo acidulante ácido cítrico e antioxidante ácido ascórbico (**Tabela 1**).

É importante ressaltar que foi feito um estudo prévio dos rótulos dos néctares comerciais, a fim de selecionar os aditivos que eram comuns à maioria dos produtos e que estavam de acordo com os aditivos previstos na RDC nº 8, de 6 de março de 2013. A resolução dispõe sobre a aprovação de uso de aditivos alimentares para produtos de frutas e de vegetais e geleia de mocotó, em que o néctar de fruta está inserido (Art 2º), e estabelece os limites máximos permitidos para cada aditivo em cada um dos produtos descritos. Assim, aos néctares formulados foram adicionados os aditivos ácido cítrico, ácido ascórbico e goma guar.

Para os aditivos ácido cítrico e ácido ascórbico, os limites máximos nos néctares de frutas são *quantum satis* (quantidade necessária para obter o efeito tecnológico desejado desde que não altere a identidade e a genuinidade do produto). Para goma guar o limite máximo estabelecido é de 0,1 g/100 mL. Para a formulação dos néctares considerou-se como limite máximo de ácido cítrico e de ácido ascórbico as maiores concentrações obtidas nos néctares comerciais (0,55 g/100 mL e 77 mg/100 mL, respectivamente). Para goma guar, considerou-se o valor correspondente a 75 % do valor máximo permitido.

O suco/polpa de cada fruta e o xarope com os aditivos foram homogeneizados (homogeneizador Fisatom 752, Perdizes, SP) e envasados em frascos de vidro âmbar de 250 mL com tampa de rosca de plástico, ambos previamente esterilizados em autoclave a 100 °C por 10 min. Após envase dos frascos com os néctares, estes foram pasteurizados em autoclave a 100 °C por 10 min (Paltrinieri, Figuerola & Rojas, 1997). Em seguida, as tampas foram hermeticamente fechadas, os frascos deixados à temperatura ambiente até resfriamento e rotulados. Após abertos para análise, os frascos foram mantidos refrigerados (4 °C a 7 °C).

Para a esterilização dos frascos, estes foram tampados com papel kraft e elástico de látex amarelo, com as tampas envolvidas em papel kraft, e colocados em autoclave à 125 °C durante 15 min. Após a esterilização, os frascos e as tampas foram colocados em estufa a 100 °C durante 15 min. Após resfriamento, foram acondicionados em caixas até o momento de envase dos néctares.

4.5 MÉTODOS DE ANÁLISE

4.5.1 Determinação dos parâmetros do PIQ

No **Capítulo I**, os métodos utilizados para determinação dos parâmetros do PIQ se basearam na IN nº24 de 8 de janeiro de 2005 (MAPA, 2005). Para determinar o pH dos néctares foi utilizado um potenciômetro que permite a determinação direta, simples e precisa do pH. A acidez total foi determinada por titulação potenciométrica utilizando soluções de álcali padrão. Para determinar a concentração de sólidos solúveis dos néctares foi utilizado o índice de refração medido por refratômetro. Para determinação de açúcares totais foi utilizado o método Eynon Lane por titulometria. A determinação do ácido ascórbico foi realizada por titulometria utilizando solução de 2,6-diclorofenolindofenol-sódio.

Nos **Capítulo II, III e IV**, os graus Brix dos néctares produzidos foram medidos em refratômetro (Hanna Instruments Brasil, Barueri, SP, Brasil).

4.5.2 Análise por espectroscopia ATR-FTIR

No **Capítulo I**, as análises feitas utilizando espectrofotômetro IV 1 foram aplicadas no experimento de diferenças varietais das maçãs. Todas as outras análises foram realizadas utilizando o espectrofotômetro IV 2.

No experimento de diferenças varietais das maçãs foram preparados três lotes de cada formulação de néctar de maçã, os quais foram transferidos para tubos Eppendorf de 3 mL e uma gota de amostra foi pipetada sobre o cristal de ATR. Cada amostra foi analisada em 10 leituras, com 16 varreduras, e os espectros foram gerados entre 4000 e 650 cm^{-1} (resolução de 4 cm^{-1}). As leituras foram feitas em ordem aleatória, sob condições de repetibilidade.

Os néctares de uva, laranja, pêsego e maracujá, formulados foram transferidos para tubos Falcon de 50 mL e aplicou-se aproximadamente 1,5 mL de cada amostra na superfície de ATR do espectrofotômetro IV 2. Foram feitas três leituras de cada amostra, com 16 varreduras, gerando espectros entre 4000 e 650 cm^{-1} (resolução de 4 cm^{-1}). As leituras foram feitas em ordem aleatória, sob condições de repetibilidade. Após cada medição, foi aplicada a correção atmosférica para evitar interferências de CO_2 . Após a aquisição de cada espectro, o acessório ATR foi limpo com algodão e acetona de grau analítico.

No **Capítulo II e III**, para cada amostra, três leituras foram registradas com 16 varreduras, resolução de 4 cm^{-1} , gerando espectros entre 4000 a 650 cm^{-1} . Uma correção de linha de base foi realizada após cada medição para evitar a interferência atmosférica e reduzir o ruído instrumental.

4.5.3 Análise quimiométrica

As análises multivariadas dos dados obtidos no espectrofotômetro de ATR-FTIR e no espectrômetro RMN-DT foram realizadas empregando o *software* de estatística multivariada MATLAB, versões 7.9.0.529 e 8.0.0.783 (Natick, MA, EUA), utilizando o pacote PLS Toolbox, versões 5.2.2 e 7.0.2 (EigenvectorsResearch Inc., Wenatchee, WA, EUA).

Os espectros ATR-FTIR foram gerados usando o software IR Solution (Shimadzu, Kyoto, Japão) e os dados foram armazenados em termos de valores de absorvância (A) e foram extraídos usando a extensão .txt.

4.5.4 Seleção de variáveis

No **Capítulo I**, para o método de seleção de variáveis OPS, foi utilizada a rotina disponível para *download* no site: <http://lqta.iqm.unicamp.br/portugues/Downloads.html#ops>. Os métodos iPLS e GA

foram aplicados usando a interface do PLS ToolBox (Wise *et al.*, 2006). Na construção dos modelos de calibração PLS foram utilizados os métodos de seleção de variáveis OPS, iPLS-GA e GA, a fim de selecionar regiões do espectro que apresentassem informações relevantes e que estivessem melhor correlacionadas com o teor de fruta principal contido nas amostras.

No **Capítulo III**, para a construção dos modelos de classificação para detecção da mistura de adulterantes foram utilizados os métodos de seleção de variáveis iPLS, GA e VIP scores (interface do PLS ToolBox), além da combinação de iPLS com GA e VIP, de forma a encontrar o melhor método para prever/detectar as misturas de adulterantes.

4.5.5 Análise por RMN-DT

A análise de RMN-DT foi realizada no espectrômetro SpecFIT com frequência de campo magnético de 15 MHz. O T_2 foi medido pela sequência de pulso Carr-Purcell-Meiboom-Gill (CPMG) com tempo de eco de $\tau = 500 \mu\text{s}$, 5000 ecos coletados, $195 \mu\text{s}$ para pulsos a 90° , $390 \mu\text{s}$ para pulsos a 180° e tempo de recirculação de 5 s. Antes da análise, os tubos foram colocados em banho termostático a 33°C até atingir temperatura constante. No momento da análise, os tubos foram homogeneizados, colocados no equipamento de RMN-DT na mesma temperatura do banho termostático ($33,0 \pm 0,1^\circ\text{C}$) e medidos.

CAPÍTULO I. DETERMINAÇÃO DAS CONCENTRAÇÕES DE FRUTAS PRINCIPAIS EM NÉCTARES ADULTERADOS POR ESPECTROSCOPIA POR REFLETÂNCIA TOTAL ATENUADA NO INFRAVERMELHO MÉDIO COM TRANSFORMADA DE FOURIER (ATR-FTIR) COMBINADA COM MÉTODOS DE CALIBRAÇÃO MULTIVARIADA E SELEÇÃO DE VARIÁVEIS

RESUMO

Os néctares comerciais de uva, pêsego, laranja, maracujá, caju e maçã, de sete diferentes marcas, foram analisados em relação aos padrões de identidade e qualidade (PIQ), apresentaram conformidade com os padrões mínimos regulamentados, com exceção do néctar de caju da marca E para o parâmetro ácido ascórbico. Constatou-se que algumas marcas reportaram teores próximos aos limites mínimos permitidos, o que pode ser considerado uma forma de redução de custos durante a produção dos néctares. Os néctares formulados de uva, laranja, pêsego e maracujá incluíram néctares não adulterados e adulterados por diluição com xarope, adição de suco de maçã e adição de suco de caju, isoladamente, em 10 níveis para cada adulterante. Eles apresentaram conformidade com os parâmetros do PIQ e com os néctares comerciais. A espectroscopia por ATR-FTIR foi utilizada para obter os espectros. Modelos de calibração multivariada por mínimos quadrados parciais (PLS) aliados a diferentes métodos de seleção de variáveis, como mínimos quadrados parciais por intervalo (iPLS), seleção de preditores ordenados (OPS) e algoritmo genético (GA), foram utilizados para quantificar as frutas principais. A seleção de variável iPLS-OPS apresentou a maior capacidade preditiva para quantificar os principais teores de frutas. As variáveis selecionadas nos modelos finais variaram de 72 a 100; os erros médios quadrados de predição foram estimados de 0,5 % a 2,6 %; os coeficientes de correlação da predição variaram de 0,948 a 0,990; e os erros relativos médios de predição variaram de 3,0 % a 6,7 %. Todos os modelos desenvolvidos foram validados.

Palavras-chave: Fraude em alimentos. Espectroscopia no infravermelho médio. Análise multivariada. Quimiometria. Seleção de variáveis. Validação multivariada.

I.1 PRODUÇÃO

Miaw, C. S. W.; Assis, C.; Silva, A. R. C. S.; Cunha, M. L.; Sena, M. M.; de Souza, S. V. C. Determination of main fruits in adulterated nectars by ATR-FTIR spectroscopy combined with multivariate calibration and variable selection methods. **Food Chemistry**, v. 254, p. 272-280, 2018.

Miaw, C. S. W.; Silva, A. R. C. S.; Cunha, M. L.; Sena, M. M.; de Souza, S. V. C. Detecção de adulteração em néctares de laranja e uva utilizando espectroscopia no infravermelho médio e análise quimiométrica. In: VIII Congresso Latinoamericano, XIV Congresso Brasileiro de Higienistas de alimentos e VI Encontro do sistema brasileiro de inspeção de produtos de origem animal, 2017, Fortaleza. **Anais...** Fortaleza: Revista Higiene Alimentar. p. 2435-2440.

Miaw, C. S. W.; Assis, C.; Silva, A. R. C. S.; Cunha, M. L.; Sena, M. M.; de Souza, S. V. C. Calibração multivariada e seleção de variáveis: determinação de frutas principais em néctares adulterados por espectroscopia ATR-FTIR. In: IX Workshop de Quimiometria, 2018, Natal. **Anais...** Natal: Livro de resumos. p. 89.

I.2 INTRODUÇÃO

Métodos quantitativos empregando técnicas espectroscópicas vibracionais requerem o uso conjunto de ferramentas quimiométricas. No entanto, os métodos de calibração multivariada para quantificar adulterações em sucos e bebidas similares são escassos na literatura científica. Dois trabalhos foram encontrados aplicando infravermelho para quantificar diferentes variedades em misturas de suco de uva (Snyder *et al.*, 2014) e para detectar adulteração de suco de laranja pela adição de xarope (Ellis *et al.*, 2016).

Outros artigos utilizaram PLS em combinação com espectros de infravermelho médio para determinar parâmetros físico-químicos de sucos de uva branca (Shah *et al.*, 2010) e em combinação com espectros de infravermelho próximo para determinar parâmetros de qualidade em goiabas congeladas e polpas de maracujá (Alamar *et al.*, 2016). Entretanto, estudos aplicando técnicas vibracionais para avaliar e quantificar adulterações em néctares de frutas pela adição de diferentes espécies ainda não haviam sido publicados.

Outro aspecto importante a ser destacado neste capítulo é a otimização do modelo por seleção de variáveis. Os métodos de seleção de variáveis são úteis para excluir variáveis irrelevantes e ruidosas, melhorando o desempenho e a robustez do método, aumentando a relação sinal-ruído e reduzindo a complexidade do modelo (Andersen & Bro, 2010). Sorol *et al.* (2010) discriminaram as estratégias de seleção de variáveis em dois tipos. O primeiro tipo é baseado na triagem contínua de vetores informativos estimados a partir de modelos multivariados. O segundo tipo é baseado em algoritmos de busca, com o objetivo de selecionar variáveis associadas a erros mínimos de previsão. O primeiro tipo de método tem a vantagem de ser mais rápido. No entanto, esses autores recomendam a complementação de um tipo de método com outro.

Os métodos de seleção de variáveis também podem ser classificados como baseados em uma busca discreta ou contínua de variáveis. No presente estudo, três métodos diferentes foram comparados. O método PLS por intervalo (iPLS)

(Norgaard *et al.*, 2000) é o método de seleção de variáveis contínuas mais comum e emprega uma estratégia do primeiro tipo. O algoritmo genético (GA) (Broadhurst *et al.*, 1997) é o método discreto mais comum e emprega uma estratégia do segundo tipo. Finalmente, a seleção de preditores ordenados (OPS) (Teofilo *et al.*, 2009) é um método discreto, mais recente e menos conhecido, que emprega um híbrido das duas estratégias de seleção de variáveis mencionadas. Combinações de iPLS com GA e com OPS também foram testadas.

Neste capítulo, foram analisados e comparados os perfis espectrais das duas variedades de maçã, Gala e Fuji, utilizadas pela indústria brasileira na fabricação de néctares, a fim de averiguar se haveria diferença entre elas para a posterior fabricação dos néctares para as análises de adulteração. Também foram caracterizados néctares de uva, pêssego, laranja, maracujá, caju e maçã de diferentes marcas comerciais em relação aos parâmetros do PIQ.

A originalidade deste estudo foi o desenvolvimento e validação de modelos de calibração multivariada PLS aliados a métodos de seleção de variáveis para quantificar os teores de frutas principais nos néctares de uva, laranja, pêssego e maracujá. Os néctares formulados incluíram amostras não adulteradas e adulteradas por diluição com xarope, suco de maçã e suco de caju. Os quatro néctares de frutas foram escolhidos por serem os néctares mais consumidos no Brasil (ABIR, 2015).

Na maioria dos artigos citados anteriormente, foram desenvolvidos métodos para análise de amostras comerciais. Como as composições dos néctares comerciais brasileiros não podem ser consideradas confiáveis em relação às quantidades mínimas de polpa, optou-se por fabricar as amostras de néctares a partir de matérias-primas confiáveis, que representassem as amostras comerciais, e construir modelos multivariados usando apenas essas amostras. Finalmente, foi realizada uma validação analítica multivariada completa para demonstrar a adequação ao propósito dos métodos desenvolvidos.

I.3 MATERIAL E MÉTODOS

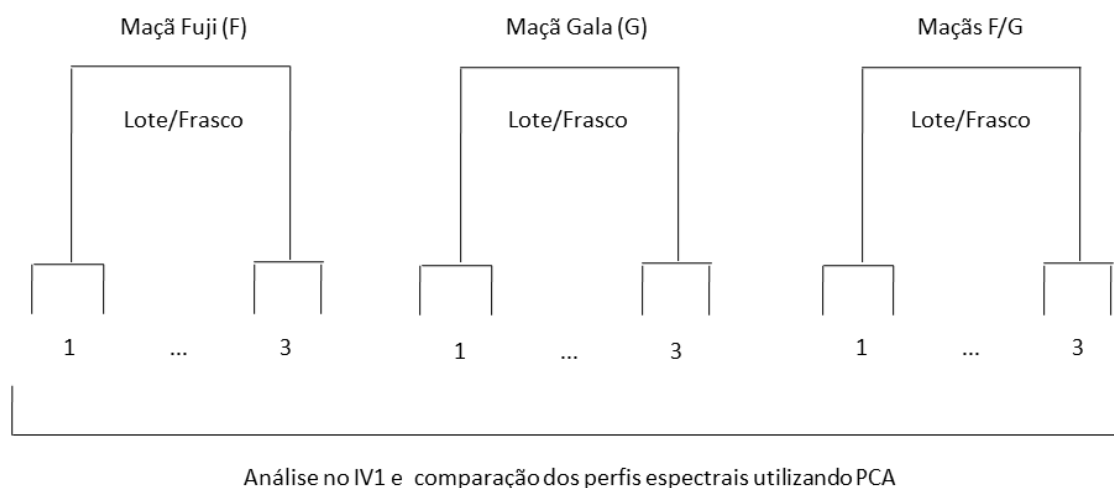
I.3.1 DELINEAMENTO EXPERIMENTAL

I.3.1.1 Diferenças varietais

De acordo com a ABIR, na formulação de néctares são utilizadas maçãs das variedades Fuji e Gala (ABIR, 2015). Então, nesta etapa do projeto, foram formulados néctares de maçã Fuji, maçã Gala e de maçã Fuji e Gala (mistura em iguais quantidades de cada variedade), em conformidade com o PIQ deste produto (30 % de suco de maçã), com o intuito de verificar se havia diferença entre os perfis espectrais das duas e definir qual delas seria utilizada nos experimentos seguintes. As formulações foram preparadas conforme o esquema da **Figura I.1**.

Os perfis espectrais foram comparados utilizando análise de componentes principais (PCA). A presença de *outliers* foi verificada por meio da análise dos valores de T^2 de Hotelling e dos resíduos Q das amostras.

Figura I.1 Delineamento do experimento de diferenças varietais



I.3.1.2 PIQ dos néctares comerciais

As amostras de néctares comerciais foram analisadas, em duplicata, para avaliação dos parâmetros do PIQ acidez total, °Brix, açúcares totais e ácido ascórbico. Foram avaliadas as médias, desvios e amplitudes destes parâmetros para os néctares das diferentes frutas. As amplitudes estimadas para os néctares comerciais, excluídos aqueles que porventura estivessem fora dos padrões regulamentados, foram utilizadas como base para os néctares formulados.

I.3.1.3 Néctares formulados

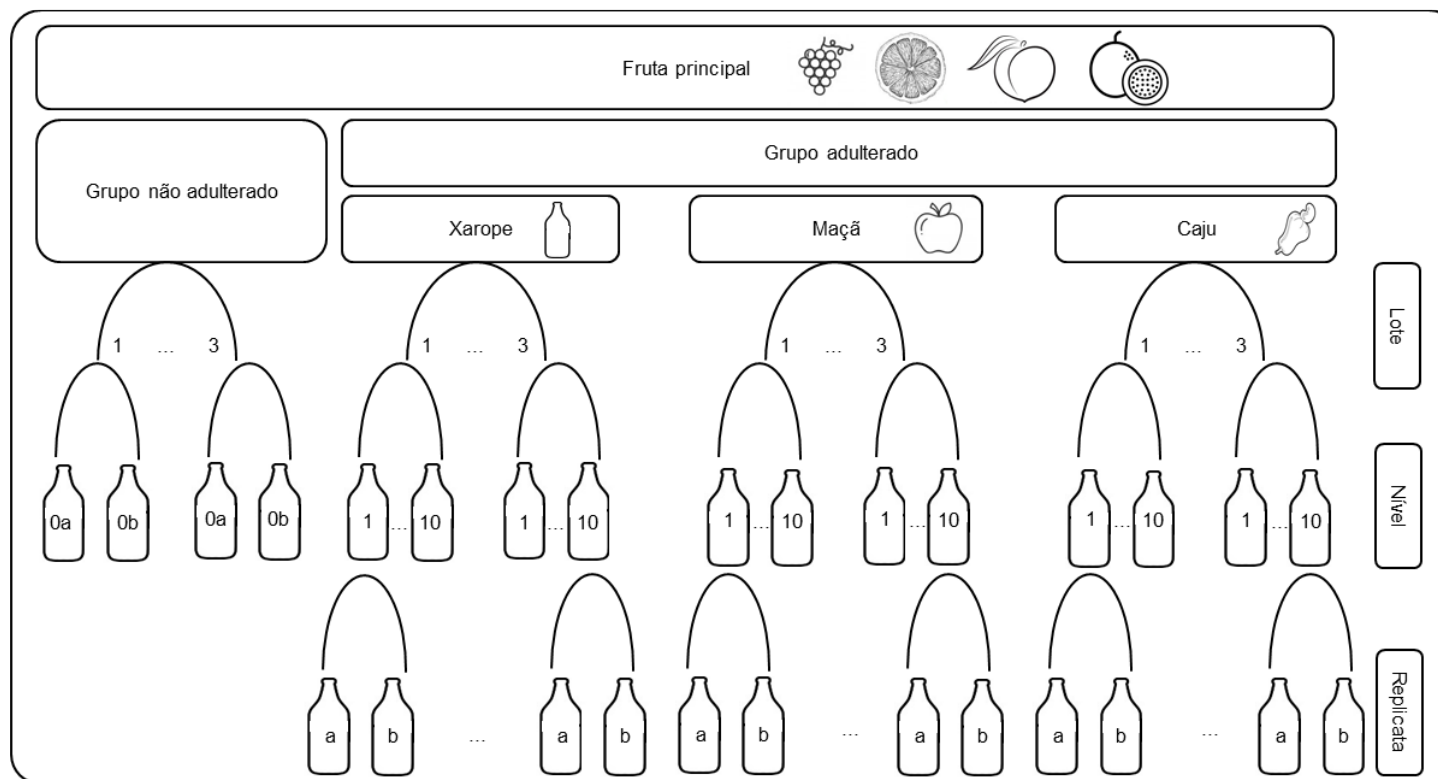
Néctares de quatro frutas, correspondentes aos mais consumidos no Brasil, a saber, uva, pêsego, laranja e maracujá (ABIR, 2015), foram formulados sem adulteração, adulterados por diluição com xarope, por adição de suco de maçã (variedade Fuji), e por adição de suco de caju.

As formulações avaliadas incluíram quatro grupos:

- i) grupo 1 - néctares não adulterados - contendo fruta principal (100 % do PIQ) + xarope;
- ii) grupo 2 - néctares adulterados pela diluição com xarope - contendo fruta principal (90,0; 85,0; 80,0; 75,0; 70,0; 65,0; 60,0; 55,0; 50,0 e 45,0 % do PIQ) + xarope;
- iii) grupo 3 - néctares adulterados pela adição de suco de maçã - contendo fruta principal (90,0; 85,0; 80,0; 75,0; 70,0; 65,0; 60,0; 55,0; 50,0 e 45,0 % do PIQ) + suco de maçã + xarope; e
- iv) grupo 4 - néctares adulterados pela adição de suco de caju - contendo fruta principal (90,0; 85,0; 80,0; 75,0; 70,0; 65,0; 60,0; 55,0; 50,0 e 45,0 % do PIQ) + suco de caju + xarope.

Foram produzidos três lotes de amostras para cada fruta principal, totalizando 186 amostras (**Figura I.2**).

Figura I.2 Delineamento do estudo de adulterações com xarope, maçã e caju em diferentes níveis



Na formulação dos néctares de laranja, uva, pêsego e maracujá levou-se em consideração a amplitude de sólidos solúveis estimada na análise dos néctares comerciais (11 a 13 °Brix). Utilizou-se o menor valor do intervalo (11 °Brix) na fórmula da **Equação 26** para calcular as quantidades de polpas/sucos e xarope a serem empregadas nas formulações. Para cumprir com o parâmetro de acidez, foi adicionado ácido cítrico em diferentes quantidades, dependendo da fruta principal. Para laranja, uva e pêsego foram utilizados 0,25 g/100 g e para maracujá 0,125g /100 g (baseado em experimentos preliminares realizados com néctares de todos os sabores e em diferentes graus de diluição). Em relação ao ácido ascórbico, mesmo não havendo PIQ para estas frutas, considerou-se o mínimo de 15 mg/100 g (baseado no PIQ do néctar de caju) nas formulações.

Ao final das formulações, as concentrações resultantes de cada ingrediente variaram para cada sabor de néctar, conforme descrito a seguir:

- Laranja e uva: i) grupo não adulterado - fruta principal no nível do PIQ (50 %) + xarope; ii) adulterado com xarope - 45,0; 42,5; 40,0; 37,5; 35,0; 32,5; 30,0; 27,5; 25,0 e 22,5 % da fruta principal + xarope; iii) adulterado com maçã ou caju - 45,0; 42,5; 40,0; 37,5; 35,0; 32,5; 30,0; 27,5; 25,0 e 22,5 % da fruta principal + 5,0; 7,5; 10,0; 12,5; 15,0; 17,5; 20,0; 22,5; 25,0 e 27,5 % de suco de maçã ou caju + xarope.
- Pêsego: i) grupo não adulterado - fruta principal no nível do PIQ (40 %) + xarope; ii) adulterado com xarope - 36,0; 34,0; 32,0; 30,0; 28,0; 26,0; 24,0; 22,0; 20,0 e 18,0 % da fruta principal + xarope; iii) adulterado com maçã ou caju - 36,0; 34,0; 32,0; 30,0; 28,0; 26,0; 24,0; 22,0; 20,0 e 18,0 % da fruta principal + 4,0; 6,0; 8,0; 10,0; 12,0; 14,0; 16,0; 18,0; 20,0 e 22,0 % de suco de maçã ou caju + xarope.
- Maracujá: i) grupo não adulterado - fruta principal no nível do PIQ (10 %) + xarope; ii) adulterado com xarope - 9,0; 8,5; 8,0; 7,5; 7,0; 6,5; 6,0; 5,5; 5,0 e 4,5 % da fruta principal + xarope; iii) adulterado com maçã ou caju - 9,0; 8,5;

8,0; 7,5; 7,0; 6,5; 6,0; 5,5; 5,0 e 4,5 % da fruta principal + 1,0; 1,5; 2,0; 2,5; 3,0; 3,5; 4,0; 4,5; 5,0 e 5,5 % de suco de maçã ou caju + xarope.

I.3.1.4 Avaliação da conformidade dos néctares formulados

Para verificar a conformidade dos lotes produzidos, as amostras dos lotes 1 dos néctares elaborados foram analisadas, em duplicata, para comparação com os parâmetros do PIQ e com o perfil estabelecido de amostras comerciais. O teste de Tukey com nível de significância $\alpha = 0,05$ % foi aplicado e as médias foram discriminadas.

I.3.1.5 Validação quantitativa

A linearidade, faixa de trabalho, veracidade e precisão foram calculadas utilizando protocolos de validação para dados univariados (European Commission, 2002; Souza & Junqueira, 2005; Souza, Pinto & Junqueira, 2007; Souza, 2007). Os protocolos para dados multivariados foram utilizados para calcular as figuras de mérito (FOM) relacionadas ao modelo, como sensibilidade (SEN), sensibilidade analítica (γ), desvio residual de previsão (RPD), viés e limite de detecção (LOD) (Valderrama, Braga & Poppi, 2007; Botelho, Mendes & Sena, 2013).

A **Tabela I.1** apresenta a descrição de como foi feita a análise de dados de cada parâmetro da validação quantitativa.

Tabela I.1 Parâmetros de desempenho calculados na validação quantitativa

Parâmetro	Experimento/Avaliação	Análise de dados
Linearidade	Curvas de referência <i>versus</i> valores preditos	- <i>Outliers</i> investigados pelo teste de Jacknife. - Regressão linear simples com verificação prévia das premissas de normalidade (teste de Ryan & Joiner), homoscedasticidade (teste de Brown & Forsythe) e independência dos resíduos (teste de Durbin & Watson). - Significâncias da regressão e da falta de ajuste ao modelo linear avaliadas por análise de variância.
Veracidade	Razão de desempenho do desvio (RPD) RMSEC, RMSEP e faixa do erro relativo médio	Quanto maior o RPD, melhor o modelo sendo que valores acima de 2,5 são considerados satisfatórios. Verificar se houve sobreajuste do modelo pela razão RMSEP/RMSEC.
Precisão	Triplicatas de amostras, em três níveis de conteúdo de adulterante (baixo, intermediário e alto) analisadas em condições de repetibilidade e de precisão intermediária (diferentes dias e diferentes analistas)	Estimativa do desvio padrão sob condições de repetibilidade (DPR _r) e precisão intermediária (DPR _R) por análise de variância, após verificação das premissas de normalidade e homoscedasticidade.
SEN	-	Estimada a partir do NAS.
γ	Razão entre SEN e ruído instrumental	Ruído instrumental estimado através do desvio padrão combinado de um vetor contendo quinze replicatas de espectros das leituras da célula ATR vazia.
Viés	-	Somatória de todos os erros de previsão (levar em conta os respectivos sinais).
LOD	-	Divide-se 3,3 pelo valor correspondente de γ .

Fonte: European Commission, 2002; Souza & Junqueira, 2005; Souza, Pinto & Junqueira, 2007; Souza, 2007; Valderrama, Braga & Poppi, 2007;

Botelho, Mendes & Sena, 2013

I.4 RESULTADOS E DISCUSSÃO

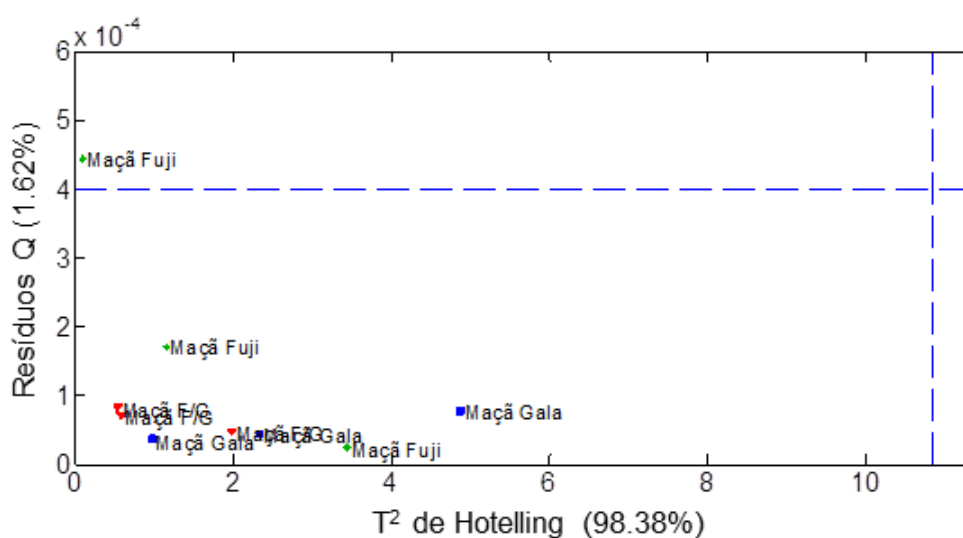
I.4.1 DIFERENÇAS VARIETAIS

No experimento buscando detectar diferenças entre variedades, os néctares de maçã foram formulados sem adulteração, em três lotes diferentes.

Uma PCA foi aplicada aos resultados obtidos após pré-processamento por MSC e centragem dos dados na média. Caso nenhuma amostra apresentasse simultaneamente altos resíduos e alta influência (quadrante superior direito), haveria indicação de ajuste do modelo. As linhas tracejadas indicam os limites de confiança a 95 % de confiabilidade desses parâmetros.

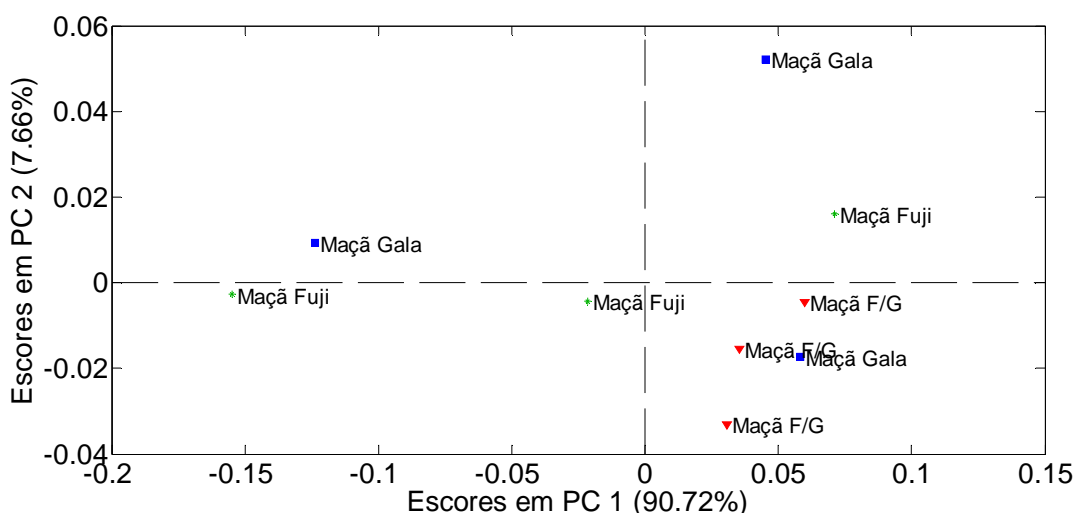
Nos néctares formulados de maçã, observando a **Figura I.3**, é possível verificar que não foram detectados *outliers*.

Figura I.3 Gráfico de valores de T^2 de *Hotelling versus* resíduos Q para detecção de *outliers* no modelo PCA construído com néctares de maçã



Os valores dos escores para as componentes principais (PC) 1 (90,72 % da variância explicada) e PC 2 (7,66 %) podem ser visualizados na **Figura I.4**, na qual se observa que não houve formação de agrupamentos, indicando que as amostras não diferiram significativamente entre si.

Figura I.4 Escores em PC1 e PC2 para o modelo construído com espectros de néctares de maçã



Com este experimento finalizado, foi possível concluir que as variedades Fuji e Gala de maçã não apresentaram diferenças espectrais significativas entre si a partir dos espectros ATR-FTIR. Assim, nos experimentos subsequentes, foi escolhida a variedade Fuji, uma vez que esta apresentou maior rendimento de suco (avaliada em experimento piloto).

I.4.2 CARACTERIZAÇÃO FÍSICO-QUÍMICA DOS NÉCTARES COMERCIAIS

As médias e desvios obtidos para os parâmetros de sólidos solúveis, açúcares totais, acidez total titulável e teor de ácido ascórbico dos néctares comerciais foram

estimadas e os valores máximo e mínimo de cada sabor foram apresentados (Tabela I.2).

Tabela I.2 Valores máximo (MAX) e mínimo (MIN), médias (MED) e desvio padrão (DP) obtidos para os parâmetros sólidos solúveis, açúcares totais, acidez total titulável e teor de ácido ascórbico dos néctares comerciais

SÓLIDOS SOLÚVEIS (°Brix)						
	CAJU	LARANJA	MAÇÃ	MARACUJÁ	PÊSSEGO	UVA
MAX	12,63	12,78	12,30	13,05	12,68	13,78
MIN	10,23	11,35	10,40	11,40	11,60	11,13
MED	11,44	12,07	11,56	12,13	11,96	12,72
DP	0,91	0,51	0,82	0,60	0,39	0,88
n	7	5	4	5	6	7
AÇÚCARES TOTAIS (%)						
	CAJU	LARANJA	MAÇÃ	MARACUJÁ	PÊSSEGO	UVA
MAX	11,60	11,10	10,60	10,70	10,90	12,20
MIN	8,70	7,00	9,20	9,20	8,90	9,40
MED	9,87	9,62	10,05	10,10	9,82	11,06
DP	0,94	1,55	0,62	0,70	0,69	1,00
n	7	5	4	5	6	7
ACIDEZ TOTAL TITULÁVEL (g/100 mL)						
	CAJU	LARANJA	MAÇÃ	MARACUJÁ	PÊSSEGO	UVA
MAX	0,24	0,53	0,44	0,55	0,34	0,49
MIN	0,15	0,39	0,23	0,40	0,22	0,27
MED	0,20	0,47	0,31	0,47	0,29	0,38
DP	0,03	0,06	0,09	0,06	0,05	0,08
n	7	5	4	5	6	7
TEOR DE ÁCIDO ASCÓRBICO (mg/100 mL)						
	CAJU	LARANJA	MAÇÃ	MARACUJÁ	PÊSSEGO	UVA
MAX	76,29	44,71	59,86	23,95	32,74	23,95
MIN	9,11	18,49	18,82	1,46	0,74	2,69
MED	42,02	30,50	33,58	12,35	16,96	13,63
DP	22,75	11,55	18,03	8,15	11,24	7,85
n	7	5	4	5	6	7

Legenda: n: número de marcas analisadas para cada sabor.

Como apenas o percentual mínimo de polpa/suco é definido no PIQ para os sabores uva, laranja e maçã, os resultados dos mesmos, mostrados na **Tabela I.2**, não foram avaliados.

Para os sabores caju, maracujá e pêssego, o teor de sólidos solúveis estava em conformidade com a legislação (mínimo de 10° Brix para caju e mínimo de 11° Brix para maracujá e pêssego). Em relação aos açúcares totais, o mínimo estabelecido é de 7 g/100 g, assim, todas as amostras de caju, maracujá e pêssego estavam em conformidade.

Em relação à acidez, todas as amostras de caju, maracujá e pêssego estavam em conformidade com o mínimo exigido no PIQ (mínimo de 0,12; 0,25 e 0,15 g/100 mL para caju, maracujá e pêssego, respectivamente).

Em relação ao teor de ácido ascórbico, este parâmetro só é definido no PIQ para néctar de caju (mínimo de 15 mg/100 mL). Assim, a marca E (validade 22/12/2016) não apresentou conformidade para o néctar de caju (9,11 mg/100 mL). A marca E também apresentou valores significativamente inferiores às demais marcas ($p < 0,05$) para os outros néctares, demonstrando uma tendência de pouca ou nenhuma adição de ácido ascórbico aos produtos.

I.4.3 CARACTERIZAÇÃO FÍSICO-QUÍMICA DOS NÉCTARES FORMULADOS

Os resultados da análise de Tukey com as médias discriminadas para os parâmetros do PIQ estão relacionados na **Tabela I.3**.

Tabela I.3 Discriminação das médias pelo teste de Tukey dos néctares formulados para os parâmetros sólidos solúveis, açúcares totais, acidez total titulável e teor de ácido ascórbico

SÓLIDOS SOLÚVEIS (°Brix)					AÇÚCARES TOTAIS (%)				
FORM	LAR	UVA	PES	MAR	FORM	LAR	UVA	PES	MAR
NA	11,75 ^b	11,60 ^{cd}	11,58 ^e	12,57 ^{bc}	NA	10,70 ^{cd}	9,70 ^{ab}	10,00 ^d	12,40 ^a
90C	11,80 ^b	11,50 ^{de}	12,38 ^a	12,73 ^{ab}	90C	11,40 ^b	9,80 ^{ab}	10,50 ^{bcd}	10,60 ^c
65C	11,50 ^c	11,53 ^d	12,10 ^{bc}	12,53 ^{bcd}	65C	10,10 ^e	9,90 ^{ab}	10,20 ^{cd}	10,90 ^c
45C	11,50 ^c	11,40 ^e	11,83 ^d	12,80 ^a	45C	10,60 ^{cde}	10,00 ^{ab}	10,90 ^{abc}	10,50 ^c
90X	11,77 ^b	11,70 ^c	11,55 ^e	12,40 ^{de}	90X	10,60 ^{cde}	9,50 ^b	10,30 ^{bcd}	11,80 ^{ab}
65X	12,45 ^a	11,87 ^b	11,75 ^d	12,47 ^{cd}	65X	11,00 ^{bc}	9,90 ^{ab}	11,00 ^{ab}	12,10 ^a
45X	12,53 ^a	11,83 ^b	12,05 ^c	12,80 ^a	45X	12,20 ^a	10,10 ^a	10,40 ^{bcd}	12,30 ^a
90M	11,48 ^c	12,00 ^a	12,25 ^{ab}	12,30 ^e	90M	10,90 ^{bc}	10,20 ^a	11,00 ^{ab}	11,20 ^{bc}
65M	11,50 ^c	12,00 ^a	12,20 ^b	12,63 ^b	65M	10,30 ^{de}	10,20 ^a	11,30 ^a	11,10 ^{bc}
45M	11,68 ^b	12,00 ^a	12,10 ^{bc}	12,60 ^{bc}	45M	10,60 ^{cde}	9,80 ^{ab}	10,10 ^d	11,20 ^{bc}
MAX	12,53	12,00	12,38	12,80	MAX	12,20	10,20	11,30	12,40
MIN	11,48	11,40	11,55	12,30	MIN	10,10	9,50	10,00	10,50

ACIDEZ TOTAL TITULÁVEL (g/100 mL)					TEOR DE ÁCIDO ASCÓRBICO (mg/100 mL)				
FORM	LAR	UVA	PES	MAR	FORM	LAR	UVA	PES	MAR
NA	0,53 ^a	0,55 ^a	0,37 ^a	0,48 ^a	NA	36,75 ^d	7,35 ^c	6,80 ^{de}	9,39 ^{def}
90C	0,49 ^c	0,51 ^b	0,36 ^a	0,44 ^b	90C	43,00 ^c	8,07 ^c	10,90 ^c	7,90 ^{ef}
65C	0,50 ^{bc}	0,47 ^c	0,36 ^a	0,36 ^c	65C	51,76 ^b	28,63 ^b	38,13 ^b	8,88 ^{ef}
45C	0,44 ^e	0,43 ^{de}	0,34 ^{cd}	0,30 ^d	45C	61,97 ^a	44,83 ^a	52,58 ^a	28,65 ^a
90X	0,51 ^b	0,52 ^b	0,36 ^{ab}	0,43 ^b	90X	34,64 ^e	4,52 ^c	4,52 ^{fg}	14,74 ^{bcd}
65X	0,44 ^e	0,45 ^{cd}	0,33 ^d	0,35 ^c	65X	28,40 ^f	9,11 ^c	9,77 ^c	11,58 ^{cde}
45X	0,37 ^g	0,38 ^f	0,30 ^e	0,29 ^d	45X	23,60 ^g	9,13 ^c	7,52 ^d	4,50 ^f
90M	0,46 ^d	0,51 ^b	0,34 ^{cd}	0,44 ^b	90M	32,74 ^e	4,64 ^c	5,80 ^{ef}	16,00 ^{bc}
65M	0,42 ^f	0,42 ^{de}	0,35 ^{bc}	0,35 ^c	65M	27,96 ^f	6,33 ^c	4,10 ^g	18,75 ^d
45M	0,42 ^f	0,40 ^{ef}	0,34 ^{cd}	0,30 ^d	45M	15,07 ^h	4,97 ^c	4,51 ^{fg}	13,22 ^{bcd}
MAX	0,53	0,55	0,37	0,48	MAX	61,97	44,83	52,58	28,65
MIN	0,37	0,38	0,30	0,29	MIN	15,07	4,52	4,10	4,50

Legenda: C: amostras adulteradas com suco de caju; LAR: laranja; M: amostras adulteradas com suco de maçã; MAR: maracujá; MAX: valores máximos; MIN: valores mínimos; NA: amostras não adulteradas; PES: pêssigo; X: amostras adulteradas com xarope. *As médias seguidas pelas mesmas letras nas colunas não diferiram entre si pelo teste de Tukey (p>0,05).

O intervalo de média dos sólidos solúveis acompanhou o dos néctares comerciais e, assim como o teor de açúcares, está de acordo com o PIQ mínimo de cada fruta principal.

O intervalo médio de acidez total titulável também acompanhou o dos néctares comerciais. Já a análise de ácido ascórbico demonstrou que os néctares que não possuem esse ingrediente originalmente em suas frutas, tiveram seu teor reduzido durante a produção ou armazenamento, indicando que o aditivo pode ter sofrido degradação. Além disso, para os néctares adulterados com caju, à medida que o teor da fruta principal (90, 65 e 45 %) diminuía, o de caju aumentava e, por consequência, o teor de ácido ascórbico aumentava, pois o caju é rico em vitamina C. De qualquer forma, os néctares foram considerados adequados e representativos dos néctares comerciais brasileiros, visto que os parâmetros analisados estavam em conformidade com os padrões estabelecidos na legislação, bem como dentro dos intervalos de amplitudes determinados para os néctares comerciais.

I.4.4. MODELO DE CALIBRAÇÃO MULTIVARIADA PARA QUANTIFICAÇÃO DAS FRUTAS PRINCIPAIS EM NÉCTARES NÃO ADULTERADOS E ADULTERADOS

A matriz de variáveis independentes (\mathbf{X}), para cada conjunto de dados de cada néctar, foi composta pelas amostras não adulteradas (fruta principal em concentração no nível do PIQ), adulteradas com caju (10 níveis), adulteradas com maçã (10 níveis) e adulteradas com xarope (10 níveis). Foram produzidas duplicatas verdadeiras de cada amostra e três lotes de cada adulteração, sendo as linhas da matriz \mathbf{X} correspondentes ao total de amostras ($n = 186$) e as colunas correspondentes às absorvâncias medidas ($n_{\text{vars}} = 1739$). O vetor com a variável dependente (\mathbf{y}) foi construído contendo a porcentagem de fruta principal nas amostras (variou de acordo com a fruta principal em questão).

O vetor \mathbf{y} possui um número de linhas igual ao número de amostras na matriz \mathbf{X} ($n=186$). Os pré-processamentos variação normal padrão (SNV), primeira derivada e

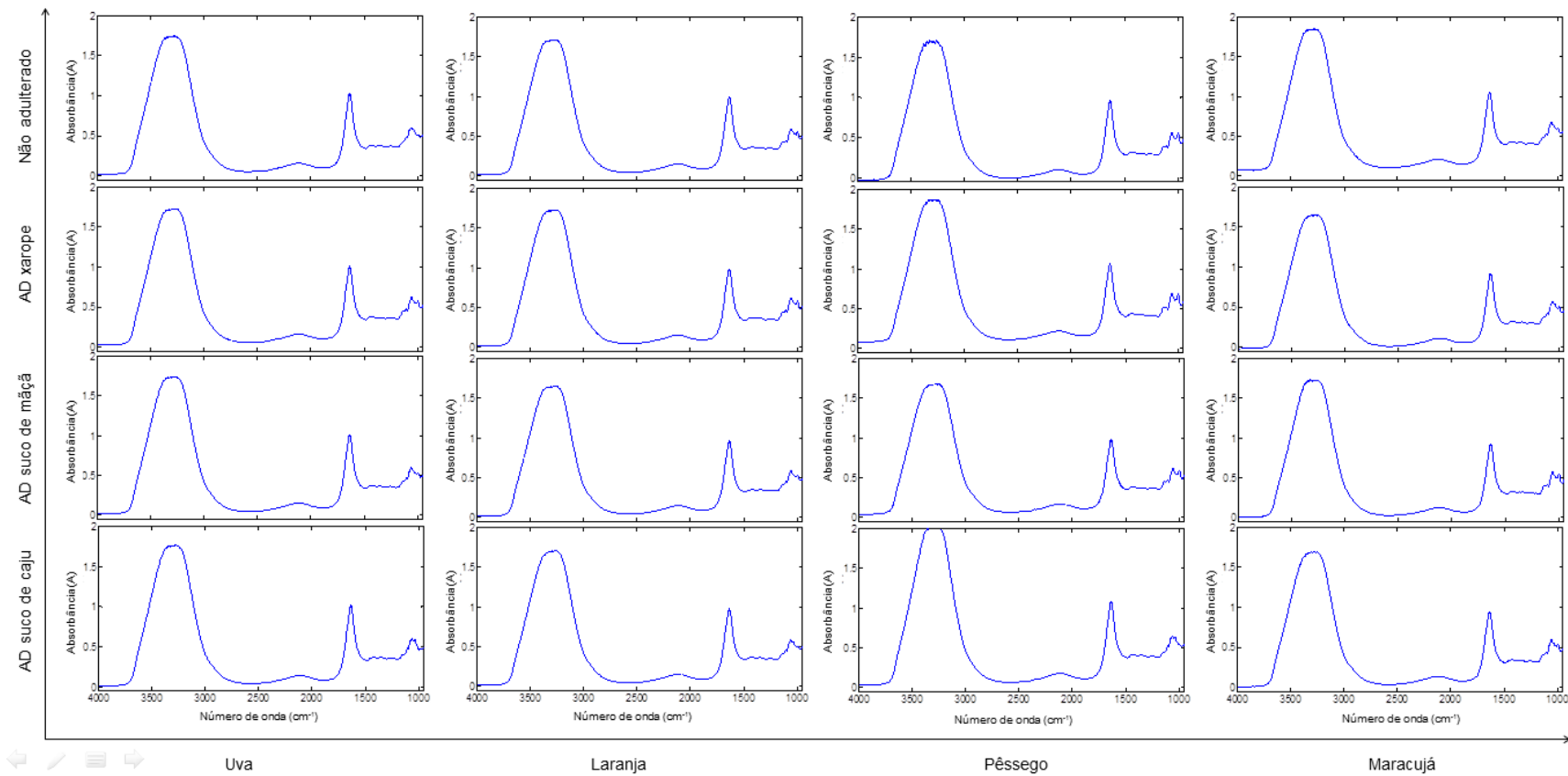
correção do espalhamento multiplicativo (MSC) foram testados, separadamente, na matriz **X**, de forma a encontrar o modelo com melhor veracidade:

A qualidade dos modelos foi avaliada comparando os erros quadráticos médios de validação cruzada e previsão (RMSECV e RMSEP) e os coeficientes de correlação de calibração (R_c), de validação cruzada (R_{cv}) e de previsão (R_p).

A detecção de *outliers* foi realizada utilizando o teste de Jackknife e aplicada nos subconjuntos de amostras de calibração e validação.

Espectros representativos obtidos no IV 1 para as amostras de néctares de uva (1), laranja (2), pêssego (3) e maracujá (4) não adulterado (a), adulterado pela diluição com xarope (b), adição de suco de maçã (c) e adição de suco de caju (d) no nível mais alto, estão apresentados na **Figura I.5**, respectivamente.

Figura I.5 Espectros representativos dos néctares de uva, laranja, pêsego e maracujá, não adulterado, adulterado pela diluição com xarope (AD xarope), adição de suco de maçã (AD suco de maçã) e adição de suco de caju (AD suco de caju), nos níveis mais altos



I.4.4.1 Atribuição espectral dos néctares formulados

Todos os espectros dos néctares avaliados apresentaram picos nas regiões de 3600 cm^{-1} a 3000 cm^{-1} e ao redor de 1640 cm^{-1} , correspondentes à absorção do grupo O-H da água (He, Rodriguez-Saona & Giusti, 2007; Shen *et al.*, 2016). Açúcares (como glicose e frutose) e ácidos orgânicos (como ácido cítrico e ácido málico) apresentaram bandas intensas e características na região de 1500 cm^{-1} a 950 cm^{-1} . Nesta região, as variações observadas estavam relacionadas ao tipo de amostra avaliada (Bureau *et al.*, 2009; Shah *et al.*, 2010; Shen *et al.*, 2016). A região de 1700 cm^{-1} a 1500 cm^{-1} corresponde ao estiramento C=O de pectinas e ácidos, mas foi sobreposta pela banda de absorção da água (Defernez & Wilson, 1995).

A região de 1500 cm^{-1} a 1200 cm^{-1} apresentou uma combinação de bandas e pode estar relacionada às deformações de CH₂, C-C-H e H-C-O (Sivakesava & Irudayaraj, 2001; Irudayaraj & Tewari, 2003; Moros *et al.*, 2005; Beullens *et al.*, 2006; Vardin *et al.*, 2008; Shah *et al.*, 2010). De acordo com He, Rodriguez-Saona & Giusti (2007), as bandas múltiplas estão relacionadas à absorção do estiramento de C-O e deformação angular de C-O-H de fenóis, ácidos carboxílicos e carboidratos (Cabaniss, Leenheer & Mcvey, 1998; He, Rodriguez-Saona & Giusti, 2007).

A região de 1200 cm^{-1} a 900 cm^{-1} é atribuída a fortes vibrações de estiramento C-C e C-O, que estão relacionadas a açúcares e ácidos orgânicos (Sivakesava & Irudayaraj, 2001; Beullens *et al.*, 2006; He, Rodriguez-Saona & Giusti, 2007; Vardin *et al.*, 2008; Shah *et al.*, 2010). O número e a posição das bandas nesta região podem mudar (dependendo da amostra), já que cada amostra apresenta diferentes proporções dos tipos de sacarídeos (Moros *et al.*, 2005). O estiramento de C-O e a deformação de OH de álcoois primário, secundário e terciário resultaram nos picos a 1150 cm^{-1} , 1100 cm^{-1} e 1050 cm^{-1} (Vardin *et al.*, 2008).

I.4.4.2 Pré-processamento dos dados

Os modelos de calibração multivariada foram criados para cada fruta principal. As regiões que apresentavam apenas ruídos experimentais foram previamente retiradas (de 936 cm^{-1} a 650 cm^{-1}), resultando em 1589 variáveis na região de 4000 cm^{-1} a 937 cm^{-1} . As técnicas de pré-processamento de sinais variação normal padrão (SNV), 1ª derivada de Savitzky-Golay e correção multiplicativa de espalhamento de luz (MSC) foram testadas, sendo os melhores resultados obtidos com o MSC, que apresentou o menor valor de RMSECV. Assim, aplicou-se aos dados o MSC seguido de centragem da média.

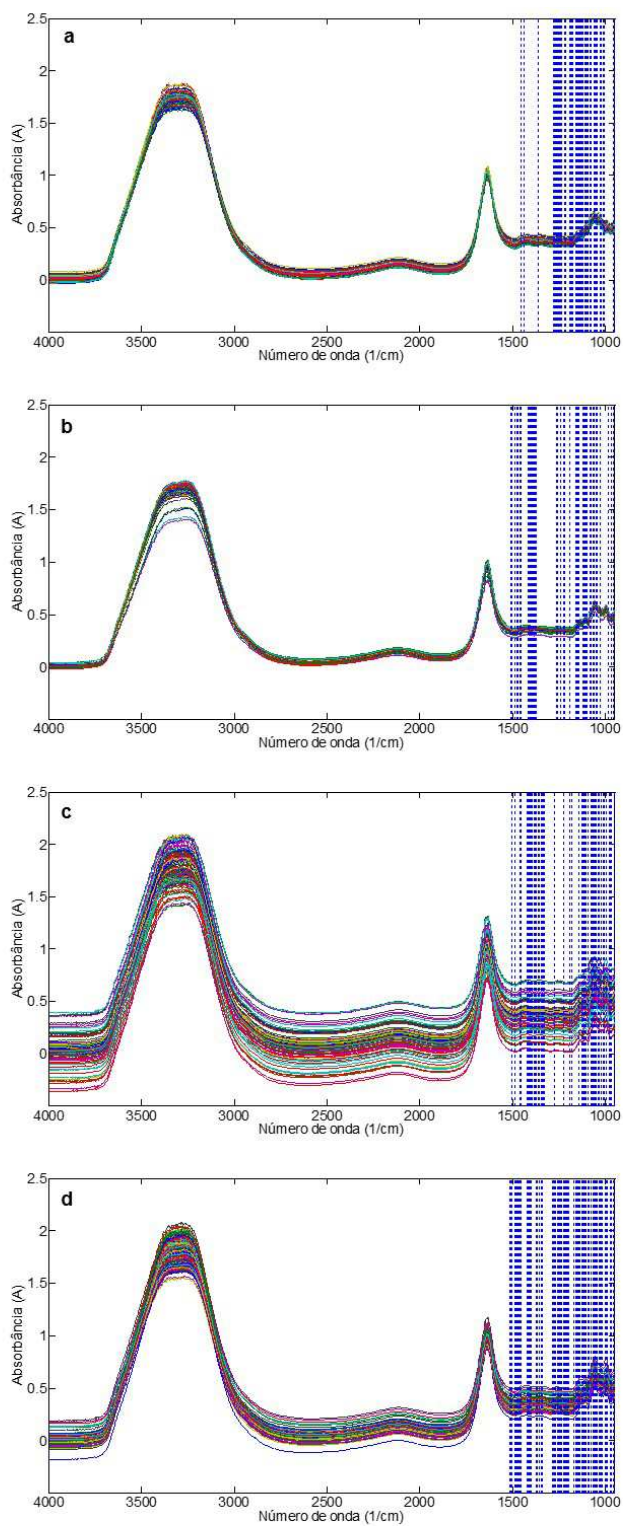
Para os modelos PLS, as amostras foram divididas em conjunto de calibração e validação. Como as amostras estavam organizadas para cada nível de concentração, 2/3 das amostras foram selecionadas para o conjunto de calibração (amostras 1, 3, 4 e 6, de cada nível, totalizando 124 amostras) e 1/3 para o conjunto de validação (amostras 2 e 5, totalizando 62 amostras). Selecionou-se o método de venezianas para a validação cruzada com 15 divisões e 20 iterações. Os *outliers* foram investigados checando as amostras que apresentavam simultaneamente altos valores de resíduos Q e altos valores de *leverage* (alavanca) bem como usando o teste de resíduos padronizados de Jackknife (Souza & Junqueira, 2005).

I.4.4.3 Modelos de calibração multivariada

Os melhores modelos PLS com espectros completos foram construídos com número de variáveis latentes (VLs) que variou de 7 a 10, dependendo da fruta principal. Esses números de VLs selecionados podem ser explicados pela alta variabilidade das amostras utilizadas nos modelos, os quais foram compostos de quatro tipos diferentes de amostras (não adulteradas e contendo 3 diferentes tipos de adulteração), além da variabilidade dos diferentes lotes (produção em diferentes dias).

Os modelos de seleção de variáveis usando iPLS, OPS, GA, iPLS-OPS e iPLS-GA foram testados com o objetivo de selecionar regiões espectrais que apresentassem informações relevantes e demonstrassem correlação com o conteúdo de fruta principal nas amostras. Em geral, todos os métodos de seleção de variáveis melhoraram os modelos. Em particular, os melhores modelos foram obtidos empregando GA e as combinações iPLS-OPS e iPLS-GA. Como um todo, os melhores resultados foram fornecidos pelo iPLS-OPS. Por esse motivo, esses modelos foram escolhidos para validação analítica. A **Figura I.6** exibe os espectros brutos no infravermelho médio de todas as amostras para os modelos de uva (a), laranja (b), pêssego (c) e maracujá (d). Em todas as figuras, as variáveis selecionadas pelo iPLS-OPS foram destacadas.

Figura I.6 Espectros completos das amostras formuladas de néctares de uva (a), laranja (b), pêssigo (c) e maracujá (d) com as variáveis destacadas selecionadas pelo iPLS-OPS



O iPLS foi inicialmente aplicado dividindo os espectros completos em 6 intervalos, cada um incluindo aproximadamente 300 variáveis. Para todos os modelos, o melhor intervalo (com RMSECV mais baixo) englobou a faixa espectral de 937 cm^{-1} a 1514 cm^{-1} . Nesta sequência, OPS e GA foram aplicados na região espectral selecionada. Os modelos iPLS-OPS foram obtidos com janelas de 30 variáveis, incrementos de 2 variáveis e a combinação de coeficientes de regressão e sinal analítico líquido (NAS) como vetores informativos. Os modelos GA (com espectro completo) foram otimizados de acordo com os seguintes parâmetros: população de 64, gerações de 100, taxa de mutação de 0,005, largura da janela de 1, convergência de 50, termos iniciais de 10 e cruzamento duplo. Para a combinação com iPLS (iPLS-GA), a única modificação foi o número de termos na população inicial, que foi definido como 15.

Os resultados para os modelos individuais iPLS e OPS não foram apresentados, já que cada um apresentou desempenho pior em comparação com as combinações com iPLS e GA. A melhoria dos modelos de calibração multivariada que combina seleção de variável por um método contínuo com a aplicação subsequente de um método discreto já foi observada na literatura. Leardi & Norgaard relataram bons resultados combinando iPLS retroativo (biPLS) e GA (Leardi & Norgaard, 2004) e Sorol *et al.* sugeriram complementar os resultados do método de seleção de variáveis contínua com algum algoritmo para uma pesquisa de janela discreta (Sorol *et al.*, 2010).

Os principais parâmetros estimados para comparar os modelos com espectro completo, iPLS-OPS, iPLS-GA e GA estão resumidos na **Tabela I.4**.

Tabela I.4 Comparação dos parâmetros estatísticos estimados para os modelos com espectro completo e com métodos de seleção de variáveis

Parâmetro	Modelo néctar de uva 22,5 - 50,0 % de uva				Modelo néctar de laranja 22,5 - 50,0 % de laranja				Modelo néctar de pêsego 18,0 - 40,0 % de pêsego				Modelo néctar de maracujá 4,5 - 10,0 % de maracujá			
	COMP	iPLS- OPS	iPLS- GA	GA	COMP	iPLS- OPS	iPLS- GA	GA	COMP	iPLS- OPS	iPLS- GA	GA	COMP	iPLS- OPS	iPLS- GA	GA
n_{VL}	7	7	6	9	7	7	6	7	10	8	8	10	10	8	9	10
n_{vars}	1589	90	49	100	1589	76	28	153	1589	72	41	112	1589	100	67	113
RPD_{cal}	1,85	4,87	5,38	4,61	2,58	4,04	4,72	3,35	1,70	3,37	3,16	3,11	1,49	3,34	3,06	3,38
RPD_{val}	1,33	6,53	5,43	6,41	2,41	3,05	2,42	3,94	1,29	3,35	2,58	2,44	1,53	2,74	2,79	3,94
RMSEC (%)	2,5	1,1	1,2	1,2	1,9	1,7	1,4	1,7	1,4	1,2	1,6	1,2	0,3	0,3	0,3	0,3
RMSECV (%)	3,6	1,6	1,4	1,7	2,8	1,9	1,7	2,2	3,4	1,8	1,9	1,9	1,0	0,4	0,5	0,5
R_c	0,934	0,989	0,988	0,987	0,967	0,976	0,984	0,974	0,972	0,980	0,952	0,979	0,979	0,978	0,976	0,975
R_{cv}	0,862	0,979	0,983	0,977	0,927	0,970	0,978	0,957	0,819	0,957	0,965	0,949	0,756	0,956	0,948	0,957
RMSEP (%)	5,0	1,2	1,4	1,3	3,3	2,6	3,4	1,9	3,7	1,8	2,3	2,3	0,8	0,5	0,5	0,4
R_p	0,774	0,990	0,984	0,990	0,912	0,948	0,912	0,967	0,808	0,959	0,931	0,928	0,868	0,954	0,947	0,970
ER (%)	12,8	3,0	3,5	3,1	8,2	6,7	8,5	5,1	11,8	5,6	7,4	7,3	10,0	5,9	6,4	5,2

Legenda: COMP: completo; ER: erro relativo de predição; n_{LV} : número de variáveis latentes; n_{vars} : número de variáveis do modelo; R_c , R_{cv} e R_p : coeficiente de correlação de calibração, de validação cruzada e de predição; RMSEC, RMSECV e RMSEP: raiz quadrada dos erros médios quadráticos de calibração, de validação cruzada e de predição; RPD_{cal} e RPD_{val} : relação de desempenho do desvio das amostras de calibração e de validação.

Para laranja e maracujá, os modelos GA apresentaram os RMSEP mais baixos e os erros relativos médios de predição (ER), mas exigiram um maior número de variáveis que o iPLS-OPS. Para uva e pêssego, o iPLS-OPS forneceu valores mais baixos RMSEP e ER. Em geral, não há diferenças significativas entre os resultados desses dois métodos, que também apresentaram os valores mais altos de coeficientes de correlação e RPD. No entanto, uma vantagem adicional do iPLS-OPS é um tempo significativamente menor de processamento de dados (alguns minutos *versus* algumas horas para o GA).

Como pode ser observado na **Tabela I.4**, o número de VLS usado para construir os modelos iPLS-OPS variou de 7 (uva e laranja) a 8 (pêssego e maracujá), representando pelo menos 98,0 % de variância no bloco **X** e pelo menos 95,3 % no bloco **Y**. Em comparação com o modelo de espectro completo, o iPLS-OPS diminuiu o valor de RMSEP, que é o principal parâmetro para avaliar a qualidade/veracidade do modelo, de 5,0 % a 1,2 % (uva), de 3,3 % para 2,6 % (laranja), de 3,7 % para 1,8 % (pêssego) e de 0,8 % a 0,5 % (maracujá). Estes valores correspondem a uma redução relativa nos erros de predição de 76, 21, 51 e 37 %, respectivamente. O número de variáveis utilizadas para a construção dos modelos iPLS-OPS foi reduzido para menos 7 % dos espectros completos originais, de 1589 para 90 (uva), 76 (laranja), 72 (pêssego) e 100 (maracujá).

É importante notar que as variáveis mais preditivas estão concentradas na região de *fingerprinting* do infravermelho (**Figura I.6**). As regiões espectrais de forte absorção de água foram deixadas fora dos modelos. De fato, todas as variáveis selecionadas estão entre 937 cm^{-1} e 1514 cm^{-1} , o intervalo original selecionado pelo iPLS. Esta região apresenta especificamente absorções atribuídas a açúcares e ácidos orgânicos, componentes importantes das bebidas de frutas analisadas (He, Rodriguez-Saona & Giusti, 2007; Vardin *et al.*, 2008; Shah *et al.*, 2010). Com base nesses resultados, pode-se inferir que a composição de açúcares e ácidos orgânicos é discriminatória entre os néctares de fruta analisados.

I.4.4.4 Validação dos modelos iPLS-OPS para néctares de laranja, uva, pêssego e maracujá

Os modelos iPLS-OPS foram validados para todos os néctares. As figuras de mérito faixa de trabalho, linearidade, SEN, γ , viés do modelo, veracidade e precisão foram calculadas, além dos parâmetros apresentados na **Tabela I.4**. Os parâmetros de FOM estimados estão na **Tabela I.5**.

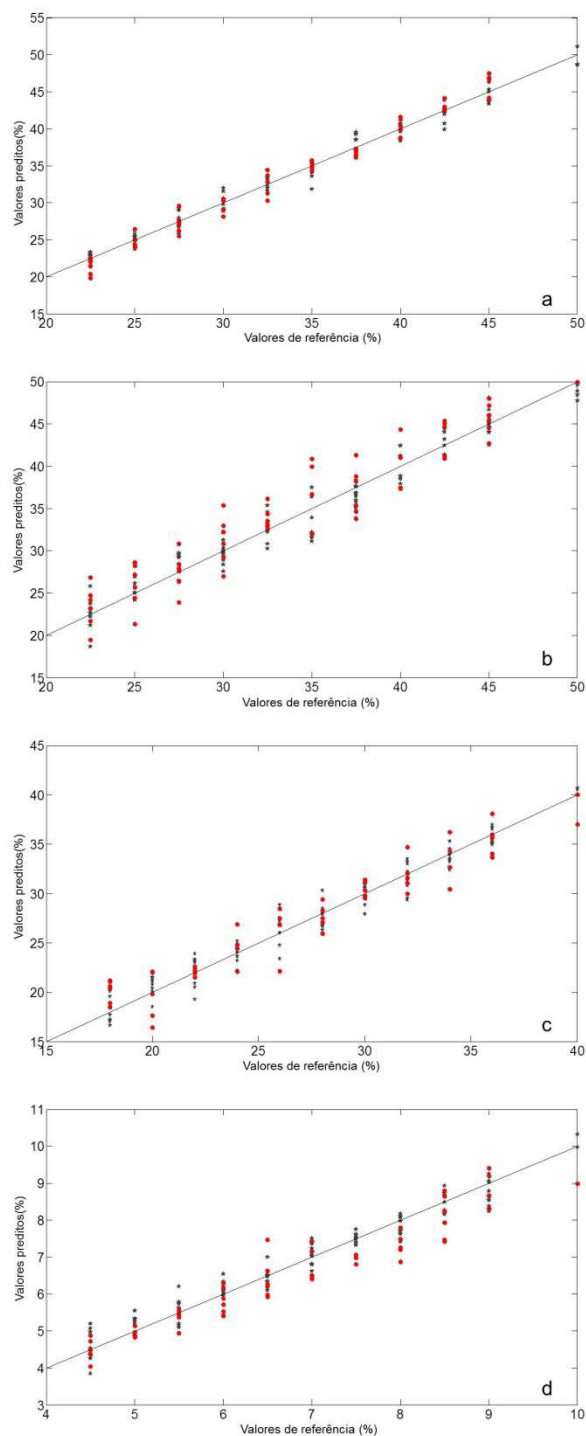
Tabela I.5 Figuras de mérito dos modelos obtidos com iPLS-OPS para néctares de uva, laranja, pêssego e maracujá

Parâmetro	Néctar de uva	Néctar de laranja	Néctar de pêssego	Néctar de maracujá	
Faixa de trabalho (%)	22,5 - 50,0	22,5 - 50,0	18,0 - 40,0	4,5 - 10,0	
Normalidade - R_{RJ} *	0,995	0,996	0,994	0,997	
Independência - D *	2,04	1,57	2,17	1,73	
Homoscedasticidade - t_L *	2,95	0,85	1,04	0,06	
Linearidade					
Regressão - F **	4537,16	1942,06	2291,15	2283,53	
Falta de ajuste - F*	0,96	2,12	0,51	1,45	
Inclinação	0,69 ± 0,50	1,60 ± 0,76	1,08 ± 0,56	0,30 ± 0,14	
Intercepto	0,98 ± 0,01	0,96 ± 0,02	0,96 ± 0,02	0,97 ± 0,02	
R^2	0,979	0,953	0,960	0,956	
SEN	$4,0 \times 10^{-3}$	$3,6 \times 10^{-3}$	$1,7 \times 10^{-3}$	$7,6 \times 10^{-3}$	
γ (% ⁻¹)	3,6	3,3	1,6	6,9	
γ^{-1} (%)	0,3	0,3	0,6	0,1	
Viés ± DP (%)	0,0 ± 1,1	0,0 ± 1,7	0,0 ± 1,2	0,0 ± 0,3	
Precisão (%)***					
DPR _r	2,4/3,9/0,2	3,5/7,2/1,9	7,1/2,9/-	4,7/2,1/0,0	
DPR _R	2,6/4,3/4,0	8,9/7,6/1,9	8,2/4,4/-	9,1/4,8/2,9	
Veracidade (%)	Erro relativo médio	3,0	6,7	5,6	5,9
LOD		0,9	1,0	2,1	0,5

Legenda: D: estatística de Durbin-Watson statistic; DP: desvio padrão; DPR_r: desvio padrão relativo no nível de repetibilidade; DPR_R: desvio padrão relativo no nível de precisão intermediária; F: razão de variância; LOD: limite de detecção; R_{RJ}: coeficiente de correlação de Ryan-Joiner; R²: coeficiente de determinação; SEN: sensibilidade; t_L: estatística t de Levene; γ : sensibilidade analítica; γ^{-1} : inverso da sensibilidade analítica; * probabilidade: p > 0,05; ** probabilidade: p < 0,05; *** valores estimados para os níveis baixo, intermediário e alto.

A linearidade pode ser avaliada através dos gráficos de valores de referência em relação aos valores preditos mostrados na **Figura I.7**.

Figura I.7 Valores de referência versus valores preditos para o modelo iPLS-OPS de uva (a), laranja (b), pêssigo (c) e maracujá (d)



Os círculos (°) vermelhos se referem ao conjunto de calibração e as estrelas (*) cinzas ao conjunto de validação.

As premissas de normalidade, homoscedasticidade e independência de resíduos foram confirmadas ($p > 0,05$). Todos os modelos apresentaram coeficientes de correlação acima de 0,948, regressão significativa ($p < 0,05$) e falta de ajuste insignificante ($p > 0,05$), indicando a linearidade de todos os modelos em seus respectivos intervalos de trabalho. É importante ressaltar que a linearidade não pode ser avaliada com base somente nos coeficientes de correlação. O comportamento aleatório dos resíduos deve ser confirmado. Para os modelos da **Tabela I.5**, os testes de Ryan-Joiner, Levene e Durbin-Watson foram utilizados para verificar a normalidade, homoscedasticidade e independência dos resíduos, respectivamente, além da análise de variância (Souza & Junqueira, 2005).

A avaliação da veracidade nos métodos multivariados é realizada principalmente comparando os valores de RMSEC e RMSEP. A maior razão RMSEP/RMSEC foi obtida com o modelo para maracujá no valor de 1,66, indicando a ausência de sobreajuste nos modelos. A veracidade também pode ser avaliada através dos erros relativos, cujos valores médios variaram entre 3,0 % (néctar de uva) e 6,7 % (néctar de laranja). O RPD é uma FOM adimensional especificamente utilizada para avaliar a veracidade dos modelos de calibração multivariada em termos absolutos. Quanto maior o RPD, melhor o modelo sendo que valores acima de 2,5 são considerados satisfatórios (Williams, 2001). Todos os valores de RPD para calibração e validação apresentaram valores superiores a 2,5, variando entre 2,74 para néctar de maracujá e 6,53 para néctar de uva (**Tabela I.4**).

A precisão foi verificada pelos valores de desvio padrão relativo de repetibilidade (DPR_r) e de precisão intermediária (DPR_R) de triplicatas obtidas em três níveis de concentração (níveis baixos, intermediários e altos). Foram obtidos valores de DPR_r e DPR_R para todos os modelos inferiores a 7,2 % e 9,2 %, respectivamente. Uma vez que a exatidão (veracidade mais precisão) e a linearidade foram estabelecidas, as faixas de trabalho puderam ser definidas entre 22,5 a 50,0 % para néctares de uva e laranja, entre 18,0 e 40,0 % para néctar de pêssigo e entre 4,5 e 10,0 % para néctar de maracujá.

Uma vez que a SEN é uma FOM dependente da técnica analítica e da matriz analisada, a γ mais robusta foi calculada com base nos valores de SEN e do ruído experimental estimado (0,001). O γ^{-1} fornece a diferença mínima de concentração que pode ser distinguida pelos métodos considerando o ruído experimental aleatório como a única fonte de erro. Esse valor também limita o número de dígitos significativos usados para expressar as previsões. Uma vez que os valores de γ^{-1} variaram de 0,1 a 0,6 %, o uso de uma casa decimal foi adotado para expressar os resultados de todos os modelos. O viés do modelo foi calculado com base apenas nas amostras de validação. Para todos os modelos, os valores de viés foram iguais a 0,0 %, indicando a ausência de erros sistemáticos.

Os LODs obtidos para uva, laranja, pêssigo e maracujá foram 0,9; 1,0; 2,1 e 0,5 %, respectivamente.

Nenhum estudo foi encontrado na literatura aplicando a espectroscopia ATR-FTIR e quimiometria para quantificar as concentrações de frutas principais em néctares de frutas por adulteração com adição de diferentes espécies. Ainda assim, os resultados obtidos podem ser comparados com os dois únicos artigos que desenvolveram métodos de calibração multivariada com base na espectroscopia ATR-FTIR para determinar a composição de sucos. Snyder *et al.* quantificaram uma variedade específica de uva em misturas de suco e obtiveram RMSEP de 5,6 %, aproximadamente 4,5 vezes maior do que o RMSEP estimado para o modelo de néctar de uva (igual a 1,2 %). Além disso, esses autores construíram seus modelos para uma faixa de trabalho de 50,0 a 100,0 % de uva, que é maior do que a do presente modelo de néctar de uva (22,5 a 50,0 %) (Snyder *et al.*, 2014). Ellis *et al.* determinaram a concentração de suco de laranja adulterado com xarope e estimaram um RMSEP de 1,7 % para o conteúdo de adulterante. Esse valor é um pouco melhor do que o RMSEP estimado para o modelo de néctar de laranja (2,6 %), porém, estes autores previram a quantidade de adulterante, ao contrário do presente estudo, que quantificou os teores das frutas principais (Ellis *et al.*, 2016). Assim, pode-se afirmar que os presentes modelos são mais robustos do que os

modelos da literatura recente, uma vez que previram o teor das frutas principais na presença de diferentes adulterantes.

I.5 CONCLUSÃO

Em relação às análises físico-químicas dos néctares comerciais, todos apresentaram conformidade com os padrões mínimos regulamentados, com exceção do néctar de caju da marca E para o parâmetro ácido ascórbico. Além disso, algumas marcas apresentaram teores próximos dos limites mínimos permitidos, o que pode ser interpretado como uma tentativa de redução de custos na produção do néctar.

Os néctares formulados apresentaram conformidade com os parâmetros do PIQ e estes parâmetros apresentaram médias próximas às dos néctares comerciais.

Os modelos PLS desenvolvidos para quantificação da fruta principal em néctares de uva, laranja, pêsego e maracujá foram aperfeiçoados por métodos de seleção de variáveis. iPLS, OPS, GA e as combinações de iPLS-OPS e iPLS-GA foram testadas e os melhores modelos foram fornecidos pelo iPLS-OPS, embora o GA tenha fornecido modelos com qualidade similar, mas usando um número muito maior de variáveis e exigindo um tempo muito maior de processamento dos dados. Os modelos iPLS-OPS foram significativamente melhores do que os modelos utilizando os espectros completos, diminuindo o RMSEP entre 76 % e 21 % e usando menos de 7 % das variáveis originais. Os números de onda selecionados estavam concentrados na região de *fingerprinting* e são associados a absorções de açúcares e ácidos orgânicos. Isso sugere que o conteúdo desses componentes nos néctares pode ser considerado discriminatório entre as frutas principais e os adulterantes.

Os métodos foram considerados simples, rápidos, diretos e de custo relativamente baixo, não consumindo reagentes ou solventes e sem a geração de resíduos químicos, de acordo com os princípios da química verde. Todas as amostras analisadas foram fabricadas a partir de matérias-primas confiáveis utilizadas pela indústria de bebidas, e os modelos foram construídos de forma robusta, incorporando amostras preparadas em diferentes lotes.

A confiabilidade dos modelos finais foi verificada estimando FOM específicas através da validação analítica multivariada completa. Todos os métodos foram validados e considerados lineares, precisos, sensíveis e sem viés, demonstrando sua adequação ao propósito de uso. Estes métodos apresentam potencial para serem adotados como ferramentas analíticas visando detectar fraudes em néctares por diluição com xarope ou adição de frutas de espécies não declaradas no rótulo. Eles podem ser de interesse para consumidores, produtores e órgãos governamentais com o objetivo de garantir a autenticidade dos rótulos de néctares de fruta.

CAPÍTULO II. DETECÇÃO DE ADULTERANTES EM NÉCTARES DE UVA POR ESPECTROSCOPIA POR REFLETÂNCIA TOTAL ATENUADA NO INFRAVERMELHO MÉDIO COM TRANSFORMADA DE FOURIER (ATR-FTIR) E ESTRATÉGIAS DE CLASSIFICAÇÃO MULTIVARIADA

RESUMO

Não há dúvidas sobre a importância do controle de fraudes alimentares, uma vez que podem causar implicações na segurança alimentar e na saúde do consumidor. Com foco nas bebidas à base de frutas, alguns tipos de adulterações foram detectados com maior frequência, como a substituição por frutas mais baratas. Uma metodologia baseada na espectroscopia por ATR-FTIR e classificação multivariada foi aplicada para detectar se os néctares de uva haviam sido adulterados por substituição com suco de maçã ou suco de caju. Um total de 126 amostras foram produzidas e analisadas. Duas estratégias foram propostas: abordagem de classe única e multiclases. Modelagem independente e flexível por analogia de classe (SIMCA), análise discriminante por mínimos quadrados parciais (PLS-DA) e mínimos quadrados parciais por modelagem de densidade (PLS-DM) foram usados para construir os modelos. Na abordagem de classe única, o PLS-DA forneceu os melhores resultados (100 % de sensibilidade e seletividade) quando comparados com SIMCA e PLS-DM, que apresentaram seletividade para o conjunto de teste de 75 % e 68 %, respectivamente. Na abordagem multiclases, o PLS-DA apresentou o melhor desempenho (sensibilidade e seletividade de quase 100%), sem predições falso-negativas. Já o SIMCA, apesar de caracterizar bem a classe adulterada com maçã, não foi capaz de diferenciar amostras não adulteradas de adulteradas com caju.

Palavras-chave: **Adulteração em alimentos. Néctar de fruta. PLS-DA. SIMCA. Modelos de classe única. Modelos multiclases.**

II.1 PRODUÇÃO

Miaw, C. S. W.; Sena, M. M.; de Souza, S. V. C.; Callao, M. P.; Ruisanchez, I. Detection of adulterants in grape nectars by attenuated total reflectance Fourier-transform mid-infrared spectroscopy and multivariate classification strategies. **Food Chemistry**, v. 266, p. 254-261, 2018.

II.2 INTRODUÇÃO

Os métodos baseados em técnicas espectroscópicas são geralmente rápidos, não destrutivos, simples e requerem pouca ou nenhuma preparação de amostra. No entanto, eles têm a desvantagem de baixa seletividade. Ferramentas eficientes para testes de adulteração podem ser criadas pela combinação dessas técnicas com métodos quimiométricos multivariados, apesar disso, alguns autores ainda aplicam apenas técnicas de estatística básica univariada (El Darra *et al.*, 2017). Os métodos de classificação são particularmente adequados para a detecção de fraudes de alimentos. Eles podem ser diferenciados em métodos discriminantes e de modelagem de classes, sendo os mais comuns a análise discriminante por mínimos quadrados parciais (PLS-DA) e a modelagem independente e flexível por analogia de classes (SIMCA), respectivamente (Bevilacqua *et al.*, 2013).

A necessidade de controle da qualidade dos alimentos foi discutida em uma revisão específica da literatura sobre o desenvolvimento de um sistema eficaz de rastreabilidade de alimentos para reduzir os numerosos casos de incidentes de segurança alimentar e de fraude (Dandage, Badia-Melis & Ruiz-García, 2017). Nesse sentido, recentemente, foram publicadas revisões abordando o uso de métodos de classificação multivariada para autenticar ou detectar adulterações em alimentos (Szymańska *et al.*, 2015; Nascimento *et al.*, 2017; Callao & Ruisánchez, 2018; Esteki, Shahsavari & Simal-Gandara, 2018). Métodos multivariados de classificação têm sido aplicados com sucesso para elucidar problemas específicos de autenticidade ou adulteração em diferentes tipos de alimentos. Alguns exemplos são: vinhos (Sen & Tokatli, 2016), óleos (Georgouli, Del Rincon & Koidis, 2017), leite (Gondim *et al.*, 2017), pastas de avelã (López *et al.*, 2014), café (Bona *et al.*, 2017), cogumelos (Xu *et al.*, 2016), vinagre (Ríos-Reina *et al.*, 2017) e uísques (Martins *et al.*, 2017).

Comparativamente, a aplicação dessas técnicas a estudos envolvendo autenticação ou detecção de fraudes em frutas e bebidas à base de frutas é mais limitada. Foram desenvolvidos artigos para classificação multivariada ou modelos de

calibração empregando diferentes técnicas analíticas, tais como espectroscopia UV-VIS (Boggia *et al.*, 2013), espectrofluorometria (Ammari, Redjdal & Rutledge, 2015), ressonância magnética nuclear (Cuny *et al.*, 2008) e espectroscopia no infravermelho médio (He, Rodriguez-Saona & Giusti, 2007; Shah *et al.*, 2010; Shen *et al.*, 2016).

Neste capítulo, a matriz néctar de uva foi estudada, por ser o sabor de néctar mais consumido no Brasil. Este estudo envolveu a detecção de adulterações do néctar de uva com suco de maçã e suco de caju, isoladamente, por meio da espectroscopia por ATR-FTIR e métodos de classificação. O suco de maçã tem sido comumente usado como matéria-prima objetivando ganhos econômicos pelas indústrias de bebidas (Singhal, Kulkarni & Rege, 1997), mas agora também vem sendo usado para substituir parte do açúcar adicionado. Além disso, o caju e a maçã são frutas suspeitas de serem utilizadas para adulterações por indústrias fraudulentas, de acordo com denúncias recebidas por órgãos governamentais (Visão Oeste, 2016), justificando a importância do desenvolvimento de métodos analíticos para detectar esses adulterantes.

Nesta aplicação, duas abordagens foram propostas considerando seus diferentes propósitos: de classe única e multiclases, utilizando métodos discriminantes e de modelagem de classes. Modelos de classe única são adequados quando o objetivo é testar se uma amostra é adulterada, independentemente de qual adulterante possa estar presente (López *et al.*, 2014). Se o adulterante é conhecido, a estratégia multiclases pode ser escolhida, pois fornece informações adicionais, como atribuições múltiplas e amostras não atribuídas a nenhuma classe (Gondim *et al.*, 2017).

Nos últimos anos, críticas em relação ao uso de métodos discriminantes (como o PLS-DA) para solucionar problemas de autenticação de alimentos têm sido levantadas, já que os resultados da classificação não serão considerados confiáveis quando o modelo for utilizado para prever amostras de uma classe não treinada (Rodionova, Oliveri, & Pomerantsev, 2016; Oliveri, 2017). Por outro lado, outros autores avaliaram a combinação de PLS-DA com detecção de *outlier*, de forma a

identificar amostras de classes não treinadas com base nos valores altos de T^2 de *Hotelling* e resíduos Q (Martins *et al.*, 2017). No entanto, como os modelos de modelagem de classe são desenvolvidos usando apenas as informações sobre amostras de uma classe de cada vez, eles não são capazes de garantir a seletividade do modelo para a detecção de várias fraudes alimentares (Xu *et al.*, 2016).

Dada a relevância dessas discussões, SIMCA e PLS-DA, como os métodos mais utilizados de modelagem de classes e discriminante, respectivamente, foram aplicados na autenticação de néctares de uva. Além disso, um método de modelagem de classe única recentemente proposto, o dos mínimos quadrados parciais por modelagem de densidade (PLS-DM) (Oliveri *et al.*, 2014), também foi aplicado. Os três métodos de classificação foram comparados em termos de sensibilidade e seletividade.

II.3 MATERIAL E MÉTODOS

II.3.1 MÉTODOS

II.3.1.1 Análise dos dados multivariados

A correção do espalhamento multiplicativo (MSC) (Rinnan, Berg, & Engelsen, 2009) foi aplicada para corrigir os desvios da linha de base não lineares dos espectros. A análise de componentes principais (PCA) foi usada como uma ferramenta de análise exploratória não supervisionada para visualizar a distribuição da amostra no espaço multivariado, visando identificar qualquer agrupamento natural nas amostras que pudessem influenciar a análise multivariada subsequente e detectar possíveis *outliers*.

Em seguida, as estratégias de classe única e multiclases foram estudadas. Como métodos de classe única foram utilizados PLS-DA, SIMCA e PLS-DM, a rotina deste método foi obtida junto ao autor, Paolo Oliveri, e seu tutorial descrito em Oliveri *et al.*, 2014. Como métodos multiclases foram aplicados PLS-DA e SIMCA (rotinas contidas no PLS Toolbox). Os métodos foram comparados em termos de sensibilidade e seletividade.

II.3.2 DELINEAMENTO

II.3.2.1 Formulação dos néctares

Conforme ilustrado na **Figura II.1**, um conjunto de 42 amostras de néctar de uva foi preparado para cada uma das três classes estudadas: não adulterada, adulterada com caju e adulterada com maçã.

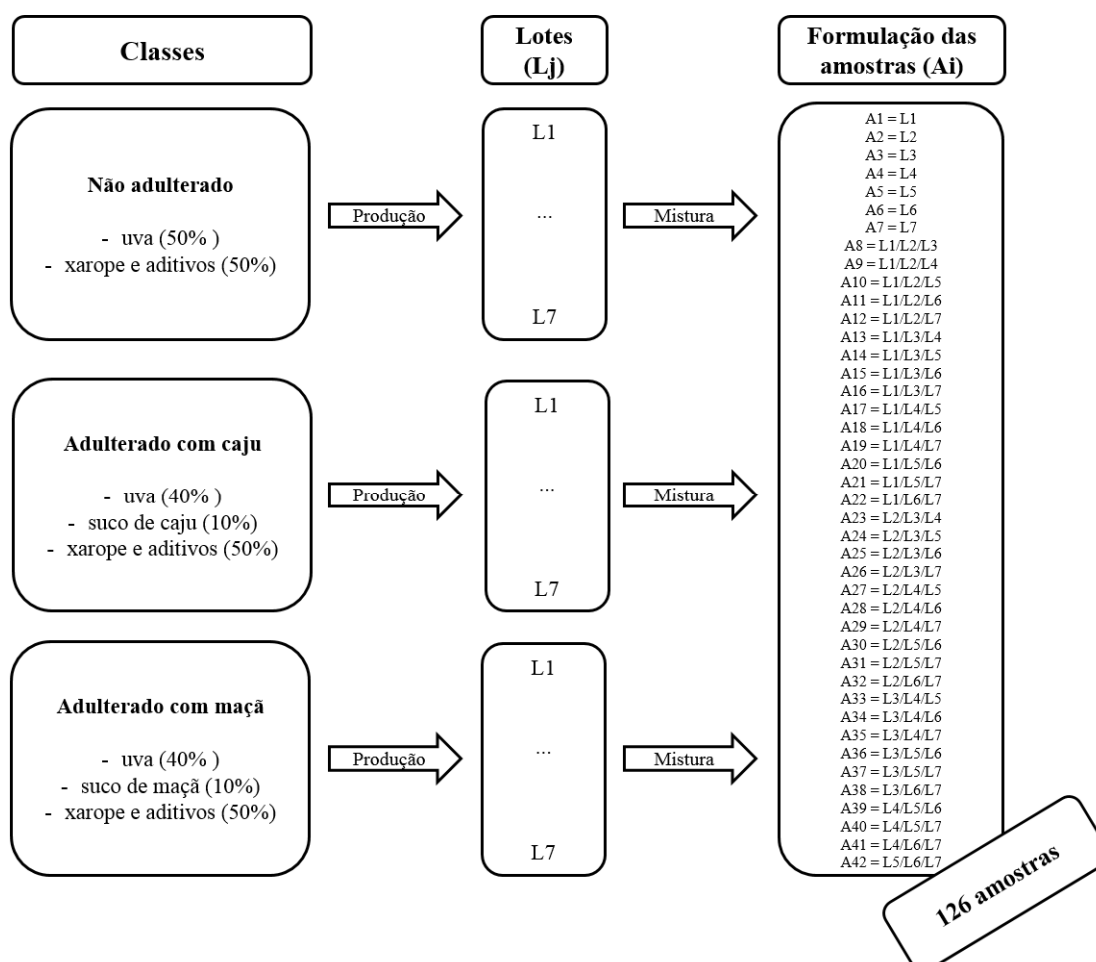
Primeiramente, sete lotes representativos de cada grupo foram preparados de acordo com as seguintes formulações:

- não adulterado (UN) - formulado com 50 % de suco de uva, 50 % de xarope de açúcar e aditivos;

- adulterado com caju (CAS) - formulado com 40 % de suco de uva, 10 % de polpa de caju, 50 % de xarope de açúcar e aditivos;

- adulterado com maçã (APP) - formulado com 40 % de suco de uva, 10 % de polpa de maçã, 50 % de xarope de açúcar e aditivos.

Figura II.1 Esquema da formulação das amostras de néctares de uva



Então, com o objetivo de obter as 42 amostras representativas de cada grupo, os 7 lotes de cada grupo foram misturados entre si em diferentes proporções, obtendo-se as 35 amostras representativas adicionais. O número final de amostras foi de 126.

O nível de concentração de adulteração de 10 % escolhido foi definido com base nos experimentos de validação quantitativa, sendo que o LOD obtido indicou que a uva, presente na forma de fruta principal, conseguiu ser detectada a níveis entre 1,0 e 50,0 %. Assim, a escolha do nível de adulteração foi baseada na concentração próxima ao LOD e considerando que concentrações muito baixas (inferiores a 10 %) não são economicamente vantajosas para a indústria fraudulenta e concentrações muito altas podem ser percebidas sensorialmente.

II.3.2.2 Parâmetros de desempenho

Os parâmetros de desempenho estimados foram sensibilidade, seletividade e a taxa de resultados inconclusivos (López, Callao & Ruisánchez, 2015; Gondim *et al.*, 2017).

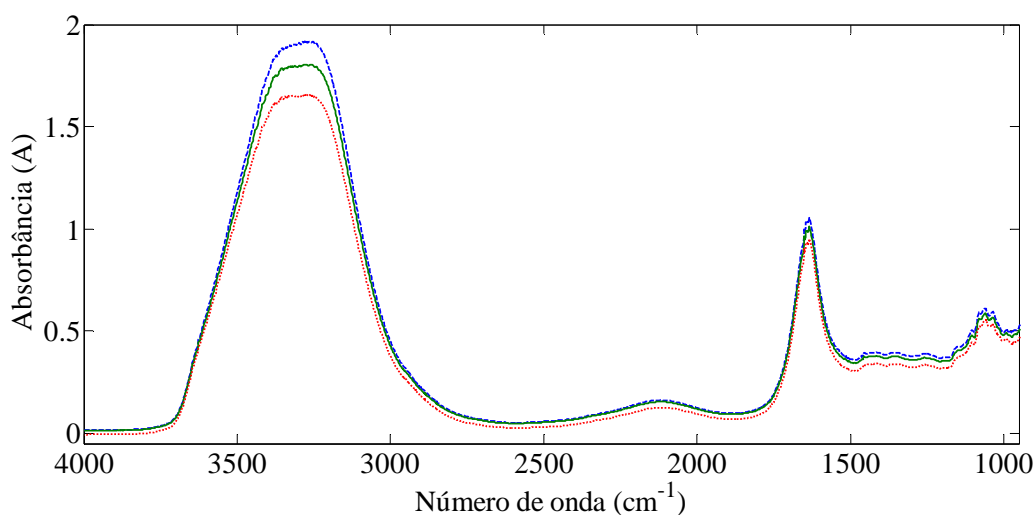
II.4 RESULTADOS E DISCUSSÃO

II.4.1 ANÁLISE ESPECTRAL

A **Figura II.2** ilustra a média dos espectros pré-processados de cada classe pré-definida em estudo. A faixa intensa próxima a 3300 cm^{-1} e o pico acentuado a 1640 cm^{-1} presentes em todas as amostras estão relacionadas à absorção O-H pela água (He, Rodriguez-Saona & Giusti, 2007; Shen *et al.*, 2016). A região entre 1700 cm^{-1} e 1000 cm^{-1} incorpora as bandas típicas para compostos fenólicos, tais como o estiramento do anel aromático C=C-C, a deformação angular de fenol OH, vibrações de deformação de C-H do anel aromático no plano e o estiramento de C-O do fenol (Coates, 2000; Bureau *et al.*, 2012). Além disso, nesta região, açúcares e ácidos orgânicos estão presentes, apresentando as bandas características (entre 1500 cm^{-1} e 950 cm^{-1}) (Shah *et al.*, 2010; Shen *et al.*, 2016). As bandas de baixa intensidade entre 1500 cm^{-1} e 1200 cm^{-1} estão relacionadas às deformações de CH₂, C-C-H e H-C-O (Vardin *et al.*, 2008; Shah *et al.*, 2010). Para a região da impressão digital (1200 a 900 cm^{-1}), as vibrações de estiramento das ligações C-C e C-O correspondem à presença de açúcares e ácidos orgânicos (He, Rodriguez-Saona & Giusti, 2007; Vardin *et al.*, 2008; Shah *et al.*, 2010).

Os componentes descritos estão presentes em todos os néctares, justificando as semelhanças entre os espectros das três classes mostrados na **Figura II.2**.

Figura II.2 Espectros médios pré-processados da classe não adulterada (linha tracejada azul), classe adulterada com caju (linha sólida verde) e classe adulterada com maçã (linha traço-ponto vermelha)



II.4.2 PRÉ-PROCESSAMENTO DOS DADOS

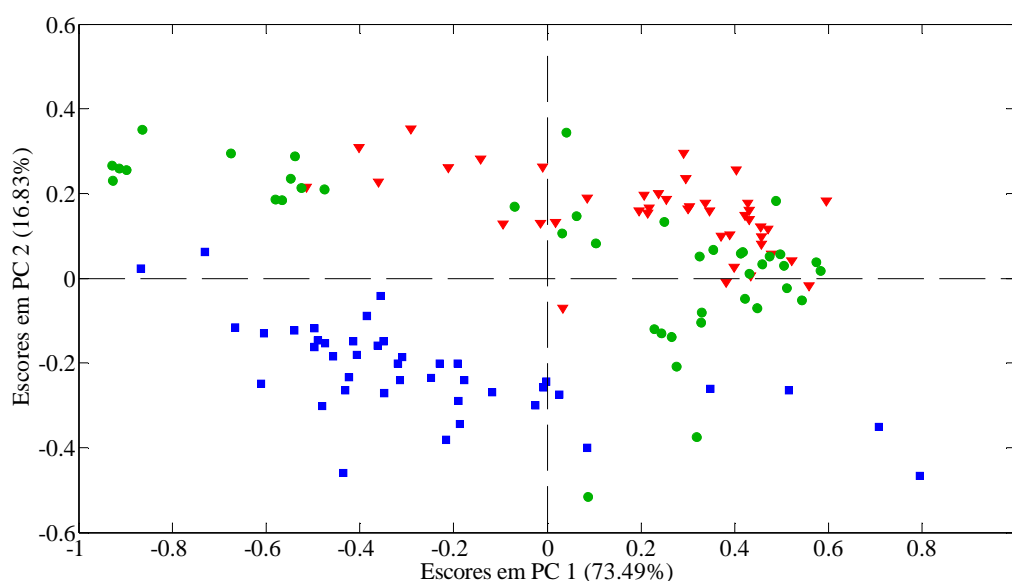
A região com variáveis muito ruidosas (de 936 cm^{-1} a 650 cm^{-1}) foram previamente removidas, resultando em 1589 variáveis na região de 4000 cm^{-1} a 937 cm^{-1} . No modelo PLS-DM, os dados foram pré-processados apenas com MSC, enquanto que nos modelos SIMCA e PLS-DA, os espectros foram tratados com MSC seguido de centragem da média.

II.4.3 ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

O gráfico de escores das duas primeiras componentes principais (PC1 x PC2), representando 90,32 % da variância total, está ilustrado na **Figura II.3**. Observa-se que a PC1 não é capaz de distinguir entre as 3 classes. Ao longo de PC2, amostras adulteradas com maçã apresentam valores de escores negativos e são claramente separadas das amostras não adulteradas, as quais apresentam valores de escores positivos. As amostras adulteradas com caju aparecem sobrepostas às amostras

não adulteradas, e apenas algumas são sobrepostas às amostras adulteradas com maçã.

Figura II.3 Escores de PC1 *versus* PC2 das amostras não-adulteradas (▼), adulterado com caju (●) e adulterado com maçã (■)



II.4.4 MODELOS MULTICLASSES

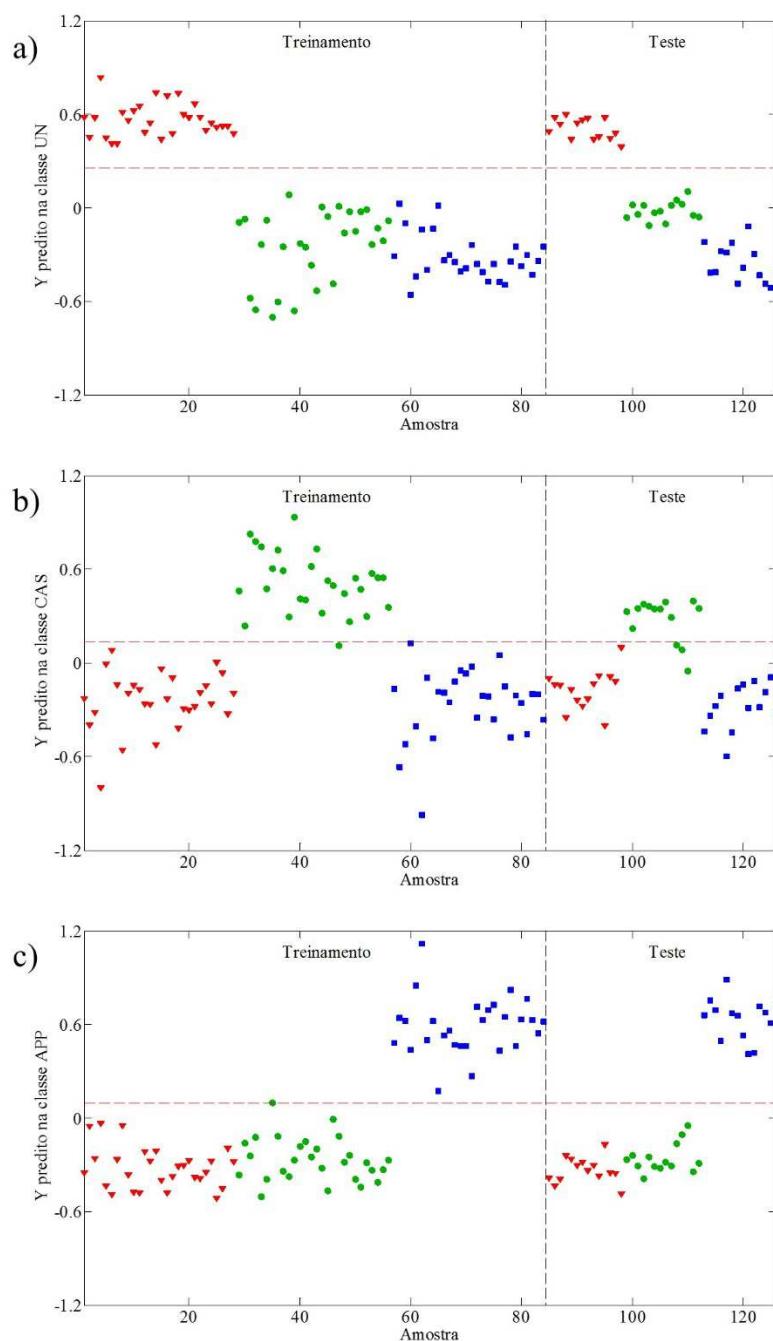
Nos modelos de classificação supervisionada, cada classe foi separada em conjuntos de treinamento e teste usando o algoritmo Kennard-Stone, sendo 28 amostras selecionadas no primeiro conjunto e 14 no segundo.

Inicialmente, implementou-se uma estratégia multiclases aplicando as técnicas de classificação SIMCA e PLS-DA para modelar as três classes pré-estabelecidas: não adulterada (UN), adulterada com caju (CAS) e adulterada com maçã (APP). Os modelos SIMCA foram construídos de forma independente para cada classe usando o conjunto de treinamento e os números ótimos de PCs foram selecionados com base no menor valor de RMSECV. Os modelos foram validados usando a validação cruzada *leave-one-out*, bem como com as previsões do conjunto teste. Foram necessárias três PCs para cada classe para construir o modelo SIMCA,

representando 95,20, 93,79 e 90,58 % da variância total, para as classes UN, CAS e APP, respectivamente.

Os modelos PLS-DA também foram construídos com as três classes. O modelo foi validado pela validação cruzada de venezianas e o número de VLS, escolhido com base no menor CVCE, foi 6, representando 95,13 % da variância no bloco **X** e 82,94 % no bloco **Y**. Os valores de *threshold* foram de 0,25 para a classe UN, 0,14 para a classe CAS e 0,09 para a classe APP, como pode ser observado na **Figura II.4**.

Figura II.4 Previsões PLS-DA para cada classe: a) não adulterada (▼), b) adulterada com caju (●) e c) adulterada com maçã (■)



Legenda: As linhas tracejadas horizontais indicam o limite de cada classe e as linhas tracejadas verticais indicam a separação entre conjunto treinamento e teste.

As atribuições de classe obtidas com os modelos SIMCA e PLS-DA estão apresentadas na **Tabela II.1**.

Tabela II.1 Previsões de classe dos modelos SIMCA e PLS-DA das amostras dos conjuntos treinamento e teste

Método	Conjunto	Classe	Classificado como						
			nºS	UN	CAS	APP	NA	MA	IR (%)
SIMCA	Treinamento	UN	28	26	28	0	0	26	92,86
		CAS	28	5	21	0	7	5	42,86
		APP	28	0	0	22	6	0	21,43
	Teste	UN	14	14	14	0	0	14	50,00
		CAS	14	7	14	0	0	7	25,00
		APP	14	0	0	14	0	0	0,00
PLS-DA	Treinamento	UN	28	28	0	0	0	0	0,00
		CAS	28	0	27	1	1	1	7,14
		APP	28	0	0	28	0	0	0,00
	Teste	UN	14	14	0	0	0	0	0,00
		CAS	14	0	11	0	3	0	21,43
		APP	14	0	0	14	0	0	0,00

Legenda: APP: classe adulterada com maçã; CAS: classe adulterada com caju; nºS: número de amostras; NA: não atribuído; MA: atribuições múltiplas; IR: resultados inconclusivos; PLS-DA: análise discriminante por mínimos quadrados parciais; SIMCA: modelagem independente e flexível por analogia de classe; UN: classe não adulterada.

Os resultados obtidos com o modelo SIMCA indicaram múltiplas classificações para as classes UN e CAS, o que já era esperado, como pôde ser observado na análise de PCA (**Figura II.3**).

Quase todas as amostras UN, tanto no conjunto treinamento como no teste, foram duplamente atribuídas à sua classe e à classe CAS. Em menor grau, amostras da classe CAS, 5 do conjunto de treinamento e 7 do conjunto de teste, também foram duplamente atribuídas à sua classe e à classe UN, e 7 das 28 amostras de treinamento não foram atribuídas a nenhuma classe. Por fim, como esperado, as

amostras da classe APP foram devidamente reconhecidas pelo seu modelo, sem atribuições erradas ou múltiplas atribuições. Apenas 6 das 28 amostras do conjunto treinamento não foram atribuídas a nenhuma classe, enquanto todas as amostras do conjunto teste foram atribuídas corretamente. Como resultado das atribuições, altas taxas de resultados inconclusivos foram obtidas para as três classes, e a classe UN foi a que apresentou a maior proporção (**Tabela II.1**).

Quanto aos resultados obtidos com PLS-DA (**Tabela II.1** e **Figura II.4**), não foram observadas atribuições errôneas. Além disso, foram observados poucos resultados inconclusivos, sendo todos correspondentes a amostras adulteradas com caju: uma amostra do conjunto de treinamento que foi duplamente atribuída, outra do conjunto treinamento e três do conjunto teste que não foram atribuídas a nenhuma classe. Deve-se enfatizar que, em nenhum caso, amostras adulteradas foram atribuídas como não adulteradas, o que significa que não foram obtidos erros falso-negativos. Do ponto de vista da fraude em alimentos, os erros falso-negativos são os mais importantes de controlar, pois correspondem aos erros relacionados à não detecção do contaminante quando ele está, de fato, presente.

II.4.5 MODELOS DE CLASSE ÚNICA

O passo seguinte foi a implementação da estratégia de classe única, na qual apenas a classe alvo foi modelada pelos três métodos de classificação. A classe UN foi considerada a classe alvo e as amostras das classes CAS e APP formaram conjuntamente a classe não-alvo. O modelo SIMCA foi construído de forma semelhante ao modelo anterior (da estratégia multiclases). A única diferença foi no cálculo da seletividade, uma vez que as amostras de CAS e APP foram modeladas juntas como classe não-alvo. O PLS-DA foi estabelecido para duas classes contrastantes, codificados como (1,0), sendo 1 para a classe UN e 0 para a classe CAS + APP. Este modelo foi construído com 5 VLs representando 94,29 % da variância no bloco **X** e 37,49 % no bloco **Y**.

Como foi explicado na seção de teoria (**item 3.3.2.3**), o PLS-DM requer a otimização de vários parâmetros: o número de k vizinhos mais próximos, de 1 a 6; tipo de pré-processamento; coeficiente de suavização a da função potencial, de 0,3 a 0,8; e o número de VLs, de 1 a 10. A etapa de otimização foi aplicada no conjunto de treinamento e, como resultado, foi obtida uma matriz de valores de sensibilidade, seletividade e eficiência (dados não apresentados) para todos os valores estudados desses parâmetros. A combinação ideal desses resultados foi avaliada considerando a maior eficiência e um número ímpar de k . Valores de k pares podem levar a classificações ambíguas, razão pela qual os números ímpares são preferidos. Os valores ótimos dos parâmetros foram definidos como $k = 3$, pré-processamento centragem na média, $a = 0,6$ e $LV = 4$.

Os resultados da classificação para esses três métodos em termos de sensibilidade e seletividade, de acordo com a estratégia de classe única, estão resumidos na **Tabela II.2**. O PLS-DA apresentou as melhores previsões, pois tanto a sensibilidade quanto a seletividade dos conjuntos de treinamento e teste foi de 100 %. Em relação aos resultados do SIMCA e do PLS-DM, eles não foram considerados satisfatórios, especialmente em relação à seletividade, uma vez que um percentual significativo de amostras adulteradas foi predito como não adulterado (25 % para SIMCA e 32 % para PLS-DM).

Tabela II.2 Sensibilidade e seletividade dos modelos SIMCA, PLS-DA e PLS-DM utilizando a estratégia de classe única

Método	Conjunto	Sensibilidade (%)	Seletividade (%)
SIMCA	Treinamento	93	91
	Teste	100	75
PLS-DA	Treinamento	100	100
	Teste	100	100
PLS-DM	Treinamento	82	91
	Teste	100	68

Legenda: SIMCA: modelagem independente e flexível por analogia de classe; PLS-DA: análise discriminante por mínimos quadrados parciais; PLS-DM: mínimos quadrados parciais por modelagem de densidade.

Quando as duas estratégias foram comparadas, pode-se afirmar que a estratégia multiclases seria preferível, pois fornece informações mais específicas sobre as adulterações. Muitas amostras na estratégia de uma classe foram atribuídas erroneamente, e na multiclases foram consideradas inconclusivas; portanto, uma análise confirmatória seria necessária.

Em relação à comparação entre os três métodos de classificação, PLS-DA, SIMCA e PLS-DM, o melhor desempenho foi claramente fornecido pelo modelo discriminante PLS-DA. Este desempenho superior de métodos discriminantes sobre métodos de modelagem de classes foi consistente com observações na literatura quimiométrica (Bylesjö *et al.*, 2006). Métodos de modelagem de classes, como o SIMCA, buscam direções de dados da maior variância, mas que podem ser distintas da direção da variação responsável pela separação de classes. Uma explicação específica para os piores resultados fornecidos pelos métodos de modelagem de classes (SIMCA e PLS-DM) neste caso foi a similaridade entre amostras das classes UN e CAS, o que foi verificado observando seus agrupamentos altamente sobrepostos no modelo PCA mostrado na **Figura II.3**.

II.5 CONCLUSÃO

A combinação de espectroscopia ATR-FTIR e técnicas de classificação permitiram a detecção de adulterações de néctares de uva com sucos de maçã e caju. Todo o procedimento analítico foi muito simples e rápido, e não exigiu o pré-tratamento da amostra ou o consumo de reagentes ou solventes. Todas as 126 amostras usadas neste estudo foram obtidas a partir de matérias-primas confiáveis e preparadas em rigorosa conformidade com as regulamentações brasileiras, exceto para as adulterações pretendidas.

Três diferentes modelos de classificação (SIMCA, PLS-DA e PLS-DM) foram desenvolvidos e duas abordagens foram consideradas: a abordagem de classe única com os três métodos e a abordagem de multiclases com SIMCA e PLS-DA. A abordagem de classe única é adequada se o principal interesse consiste em apenas detectar se a amostra é adulterada, independentemente do tipo de adulterante. Para o problema em estudo, o PLS-DA forneceu excelentes resultados, classificando todas as amostras corretamente. SIMCA e PLS-DM produziram resultados menos satisfatórios, com seletividade para o conjunto de teste de 75 % e 68 %, respectivamente.

A abordagem multiclases é a escolha apropriada quando o interesse principal é investigar a possível presença de adulterantes conhecidos. Ela fornece informações mais específicas, pois além da porcentagem de amostras atribuídas correta ou incorretamente, as informações relacionadas às atribuições inconclusivas também estão disponíveis. Na sequência, amostras inconclusivas poderiam ser submetidas à análise confirmatória. Entre os modelos multiclases, o PLS-DA também apresentou o melhor desempenho, sem predições falso-negativas, ou seja, nenhuma amostra adulterada foi classificada como não adulterada. Na análise de fraudes, é essencial evitar resultados falso-negativos, uma vez que o analista poderia declarar uma amostra como não adulterada quando na verdade é adulterada. Para a abordagem multiclases, o modelo SIMCA não foi capaz de

diferenciar amostras não adulteradas de amostras adulteradas com caju. No entanto, a classe de maçã foi bem caracterizada pelo método SIMCA.

Finalmente, podemos sugerir este tipo de aplicação como uma ferramenta potencial para auxiliar a indústria de bebidas e organismos reguladores no controle de qualidade de alimentos, permitindo a detecção em néctares de frutas por meio de análises diretas, rápidas e confiáveis.

CAPÍTULO III. SELEÇÃO DE VARIÁVEIS ALIADA À CLASSIFICAÇÃO MULTIVARIADA COM O OBJETIVO DE DETECTAR ADULTERANTES INDIVIDUAIS E SUAS MISTURAS EM NÉCTARES DE UVA

RESUMO

Durante o controle de qualidade de bebidas de frutas, alguns tipos de adulterações podem ser detectados, como a adição ou substituição por frutas mais baratas. Para determinar se os néctares de uva foram adulterados por substituição com suco de maçã ou suco de caju ou por uma mistura de ambos, uma metodologia baseada na espectroscopia por ATR-FTIR e métodos de classificação multivariada foi proposta. Os métodos de classificação análise discriminante por mínimos quadrados parciais (PLS-DA) e modelagem independente e flexível por analogia de classe (SIMCA) foram desenvolvidos como métodos multiclases (classe não adulterada, classe adulterada com caju e classe adulterada com maçã) e aplicados juntamente com diferentes métodos de seleção de variáveis. Com o intuito de melhorar a eficácia dos modelos de classificação foram testados os métodos de regressão de mínimos quadrados parciais por intervalo (iPLS), variáveis importantes na projeção dos escores (VIP escores) e algoritmo genético (GA). Os métodos de seleção variáveis melhoraram os parâmetros de desempenho para os métodos SIMCA e PLS-DA quando foram usados para prever amostras com apenas um adulterante. No entanto, apenas o PLS-DA aliado ao iPLS foi capaz de classificar amostras com misturas dos dois adulterantes, fornecendo valores de sensibilidade entre 100 e 83 % e 100 % de seletividade para as três classes estudadas.

Palavras-chave: Seleção de variáveis. Métodos multiclases. PLS-DA. SIMCA. Néctar de uva. Fraude em alimentos.

III.1 PRODUÇÃO

Miaw, C. S. W.; Sena, M. M.; de Souza, S. V. C.; Ruisanchez, I.; Callao, M. P. Variable selection for multivariate classification aiming to detect individual adulterants and their blends in grape nectars. **Talanta**, v. 190, p. 55-61, 2018.

III.2 INTRODUÇÃO

Métodos recentes foram desenvolvidos para identificar e detectar diferentes frutas em bebidas de frutas empregando várias técnicas analíticas. O uso de cromatografia líquida de ultra eficiência acoplada a espectrômetro de massas com analisador de tempo de voo (UPLC-QToFMS) (Jandrić *et al.*, 2014), cromatografia líquida de alta eficiência (Spinelli *et al.*, 2016), reação em cadeia da polimerase convencional (Han *et al.*, 2012) e em tempo real (Han *et al.*, 2012; Pardo, 2015) são alguns exemplos. No entanto, por serem técnicas mais laboriosas e dispendiosas, seu uso é mais indicado em métodos confirmatórios.

Técnicas de espectroscopia vibracional, como espectroscopia Raman e espectroscopias de infravermelho próximo e médio, são alternativas que podem ser usadas em métodos de triagem por serem mais simples, rápidas, que requerem pouco ou nenhum pré-tratamento de amostra. Os métodos desenvolvidos com base nessas técnicas estão de acordo com os princípios da química verde, sendo considerados ecologicamente corretos (Fernández-González *et al.*, 2014). Essas técnicas geram espectros que exigem análises subsequentes com ferramentas quimiométricas. Recentemente, a combinação de técnicas espectroscópicas vibracionais e métodos qualitativos multivariados proporcionou bons resultados na detecção de fraudes alimentares (Cozzolino, 2015; Callao & Ruisánchez, 2018). O uso de métodos de classificação multivariada envolve a atribuição de uma amostra a uma classe previamente estabelecida. No caso de adulterações de alimentos, uma classe corresponde ao alimento não adulterado, e diferentes classes adicionais são estabelecidas, dependendo do número de adulterantes a serem detectados. O resultado ideal atribui amostras às classes às quais elas realmente pertencem.

Como em métodos de modelos multiclasse o número de classes pré-definidas corresponde ao número de adulterantes conhecidos presente nas amostras mais uma classe (classe autêntica), os modelos multivariados desenvolvidos são comumente construídos para detectar a presença ou ausência de apenas um único adulterante por amostra. Embora seja menos comum encontrar na

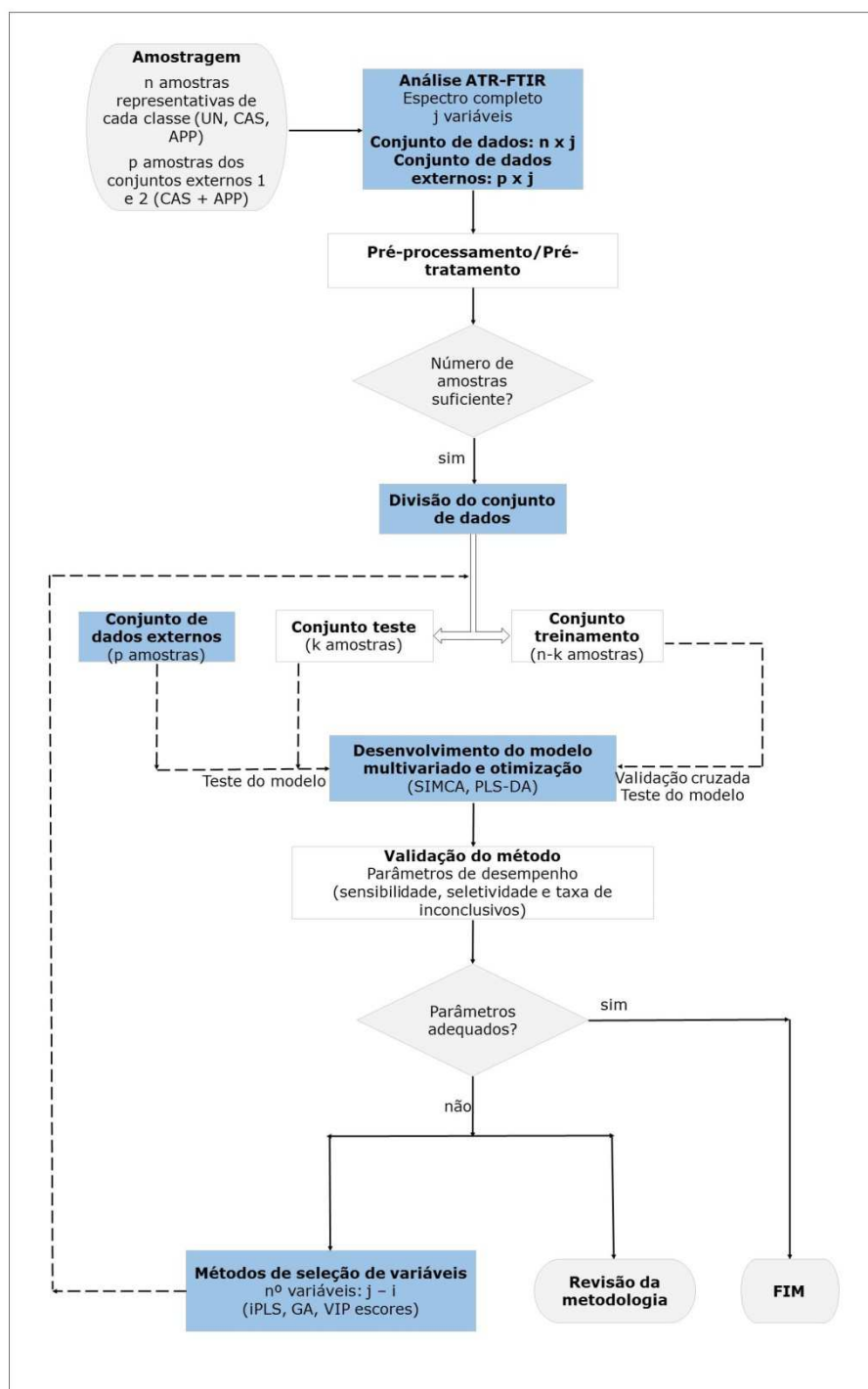
literatura estudos considerando a adulteração com misturas de dois ou mais adulterantes, esta é uma possibilidade em algumas situações reais. Portanto, é importante verificar se os métodos desenvolvidos são capazes de classificar corretamente as amostras adulteradas contendo mais de um adulterante. Estas amostras devem ser atribuídas a todas as classes estabelecidas para os respectivos adulterantes (Botelho *et al.*, 2015). Como esse problema raramente é abordado na literatura, o presente estudo representa uma contribuição importante para a comunidade científica de alimentos e pode ser facilmente estendido a outros tipos de fraudes envolvendo outros produtos ou matrizes.

A quantidade de informação gerada com a análise no infravermelho médio geralmente compreende centenas ou milhares de variáveis, e várias delas podem ser ruidosas e/ou irrelevantes para o problema em estudo. A seleção de um número limitado de variáveis/preditores informativos pode melhorar a eficácia dos modelos de classificação, reduzir sua complexidade e aumentar sua robustez (Andersen & Bro, 2010). Portanto, antes da aplicação dos métodos de classificação desenvolvidos, estratégias de seleção de variáveis podem ser empregadas na otimização. Métodos analíticos recentes têm sido encontrados na literatura que aplicam a seleção de variáveis a espectros de infravermelho com transformada de Fourier (FT-IR) em problemas de classificação (Li *et al.*, 2016; Manfredi *et al.*, 2018). Existem diferentes métodos de seleção de variáveis para remover informações ruidosas e irrelevantes, e outros que tem o objetivo de selecionar as variáveis mais discriminatórias ao trabalhar com diferentes grupos de amostras. Devido à ampla diversidade de métodos de seleção de variáveis, a escolha do mais apropriado não é uma tarefa simples (Di Anibal, Callao & Ruisánchez, 2011).

Como demonstrado no **Capítulo II**, métodos multiclases usando análise discriminante por mínimos quadrados parciais (PLS-DA) e modelagem independente e flexível por analogia de classe (SIMCA) foram previamente construídos, demonstrando bom desempenho do PLS-DA para classificação de amostras de néctares de uva contendo apenas um adulterante.

Assim, o objetivo deste capítulo foi aplicar a estratégia desenvolvida para detectar misturas de dois adulterantes em néctares de uva usando espectroscopia por ATR-FTIR. Para melhorar a capacidade destes métodos para detecção dos adulterantes maçã, caju e a mistura de ambos nos néctares de uva, três métodos de seleção de variáveis foram implementados: regressão de mínimos quadrados parciais por intervalo (iPLS) (Höskuldsson, 2001), variáveis importantes na projeção dos escores (VIP escores) (Chong & Jun, 2005) e algoritmo genético (GA) (Leardi, 2001). A estratégia proposta neste trabalho está descrita na **Figura III.1**, e demonstra esquematicamente seus diferentes passos, incluindo amostragem, metodologia de classificação (PLS-DA e SIMCA), validação analítica e estratégia de seleção de variáveis.

Figura III.1 Visão geral das etapas para desenvolvimento do método de classificação multivariada com seleção de variáveis



Legenda: APP: classe adulterada com maçã; CAS: classe adulterada com caju; CAS+APP: conjunto de amostras adulteradas com mistura de caju e maçã; UN: classe não adulterada.

III.3 MATERIAL E MÉTODOS

III.3.1 DELINEAMENTO

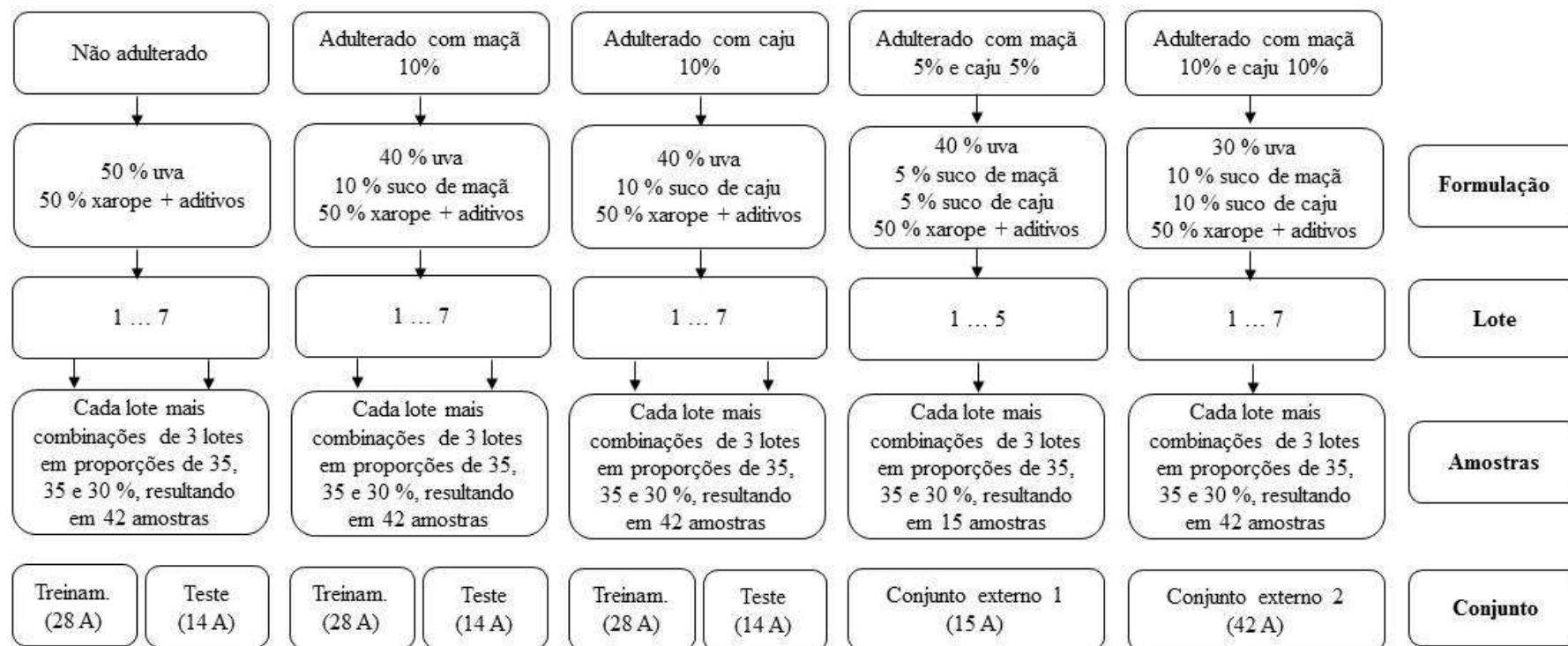
III.3.1.1 Formulação dos néctares

Adicionalmente às amostras formuladas no experimento anterior do **Capítulo II** foram formulados os conjuntos CAS + APP 1 e CAS + APP 2 compostos de misturas dos adulterantes, caju e maçã, em dois níveis de concentração diferentes:

- CAS + APP 1 - adulterado com caju e maçã - formulado com 40 % de suco de uva, 5 % de polpa de caju, 5 % de polpa de maçã, 50 % de xarope de açúcar e aditivos;
- CAS + APP 2 - adulterado com caju e maçã - formulado com 30 % de suco de uva, 10 % de polpa de caju, 10 % de polpa de maçã, 50 % de xarope de açúcar e aditivos.

A **Figura III.2** mostra um esquema do planejamento experimental para a preparação das amostras, resultando em 42 amostras representativas de cada classe, assim como 15 amostras para o conjunto externo 1 e 42 para o conjunto externo 2, totalizando 183 amostras. Espectros de cada classe foram divididos em conjuntos de treinamento (28 amostras) e de teste (14 amostras) usando o algoritmo de Kennard-Stone, o qual seleciona amostras representativas distribuídas homogeneamente no espaço multivariado.

Figura III.2 Delineamento experimental para a formulação dos néctares



Legenda: A: amostras; Treinam: treinamento.

III.3.1.2 Parâmetros de desempenho

Para avaliar a qualidade do modelo de classificação, foram calculados os parâmetros sensibilidade (SEN), seletividade (SEL) e resultados inconclusivos.

No caso particular em que as amostras adulteradas com misturas forem classificadas, deve-se ter em mente que, se ocorrerem atribuições múltiplas (para ambas as classes adulteradas – APP e CAS), a amostra não será considerada inconclusiva.

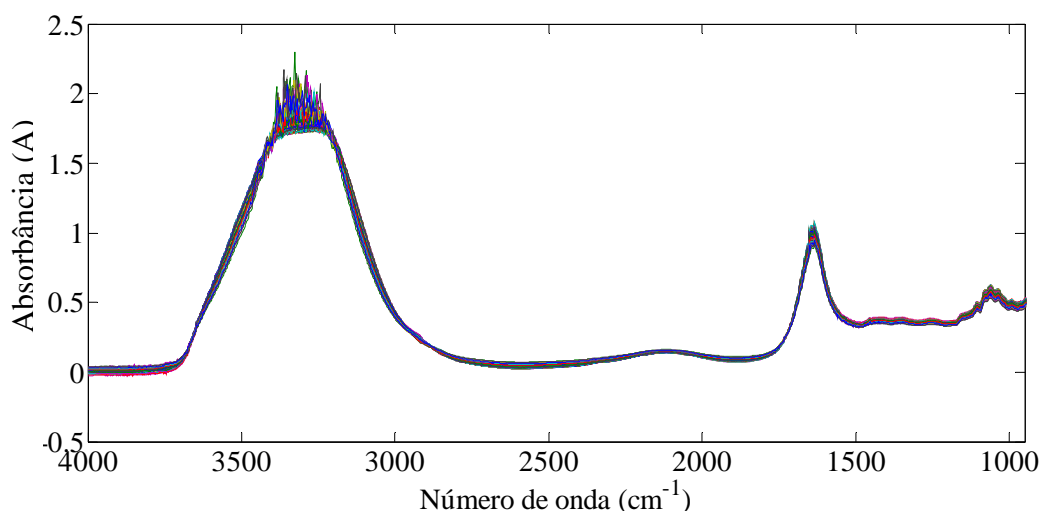
III.4 RESULTADOS E DISCUSSÃO

III.4.1 ANÁLISE ESPECTRAL

Os espectros ATR-FTIR podem ser divididos em quatro regiões, que estão relacionadas aos estiramentos de OH e CH (4000 cm^{-1} a 2500 cm^{-1}), ligações triplas (2500 cm^{-1} a 2000 cm^{-1}), ligações duplas (2000 cm^{-1} a 1500 cm^{-1}) e à região de impressão digital (1500 cm^{-1} a 400 cm^{-1}). Deformações angulares e vibrações do esqueleto caracterizam a região de impressão digital. As bandas associadas às vibrações do esqueleto normalmente refletem um padrão da molécula como um todo e não apenas um grupo específico dentro da molécula. Assim, a análise de uma impressão digital molecular vibracional pode aumentar o potencial para detectar adulterações (Stuart, 2004; Kelly & Downey, 2005).

Espectros pré-processados de todas as amostras analisadas neste estudo estão ilustrados na **Figura III.3**.

Figura III.3 Espectros de todas as amostras de néctares de uva analisadas após pré-processamento com correção do espalhamento multiplicativo



As principais vibrações dos espectros ATR-FTIR de néctares de frutas já foram atribuídas nos **Capítulos I e II**, incluindo diferentes néctares de frutas (uva, laranja, pêssigo e maracujá) e destacando néctares de uva em particular. Os sinais mais intensos nestes espectros são uma banda larga centrada em 3300 cm^{-1} e um pico acentuado em, aproximadamente 1640 cm^{-1} , ambos atribuídos às vibrações O-H da água. No entanto, a região espectral mais rica em informação parece ser a região da impressão digital, principalmente abaixo de 1200 cm^{-1} , onde se observam deformações angulares de ligações C-H, estiramentos de C-C e C-O e deformações angulares associadas a fenóis, ácidos carboxílicos e carboidratos.

III.4.2 PRÉ-PROCESSAMENTO DOS DADOS

A região espectral de 936 cm^{-1} a 650 cm^{-1} foi removida, pois apresentava muitas variáveis ruidosas. Os espectros foram pré-processados por MSC para eliminar desvios e tendências não lineares da linha de base e, em seguida, centrados na média.

III.4.3 PREDIÇÕES COM ESPECTRO COMPLETO

Os modelos multiclases SIMCA e PLS-DA utilizados para predição das amostras dos conjuntos externos 1 e 2 foram obtidos no **Capítulo II**. Estes resultados estão apresentados na **Tabela III.1** e destacados na cor cinza. Também estão apresentados os resultados de previsão dos dois conjuntos de dados externos.

Tabela III.1 Predições dos métodos SIMCA e PLS-DA utilizando o espectro completo

Método	Conjunto	Classe	Classificado como					
			UN	CAS	APP	NA	MA	IR
SIMCA	Treinamento	UN	26	28	0	0	26	26
		CAS	5	21	0	7	5	12
		APP	0	0	22	6	0	6
	Teste	UN	14	14	0	0	14	14
		CAS	7	14	0	0	7	7
		APP	0	0	14	0	0	0
	CAS + APP	1	0	0	0	15	0	15
		2	0	38	1	3	0	3
	PLS-DA	Treinamento	UN	28	0	0	0	0
CAS			0	27	1	1	1	2
APP			0	0	28	0	0	0
Teste		UN	14	0	0	0	0	0
		CAS	0	11	0	3	0	3
		APP	0	0	14	0	0	0
CAS + APP		1	1	11	7	1	5	1
		2	0	42	1	0	1	0

Legenda: APP: classe adulterada com maçã; CAS: classe adulterada com caju; CAS + APP 1: adulterado com 5 % de maçã e 5 % de caju; CAS + APP 2: adulterado com 10 % de maçã e 10 % de caju; IR: resultados inconclusivos; NA: não atribuído; MA: múltiplas atribuições; UN: classe não adulterada.

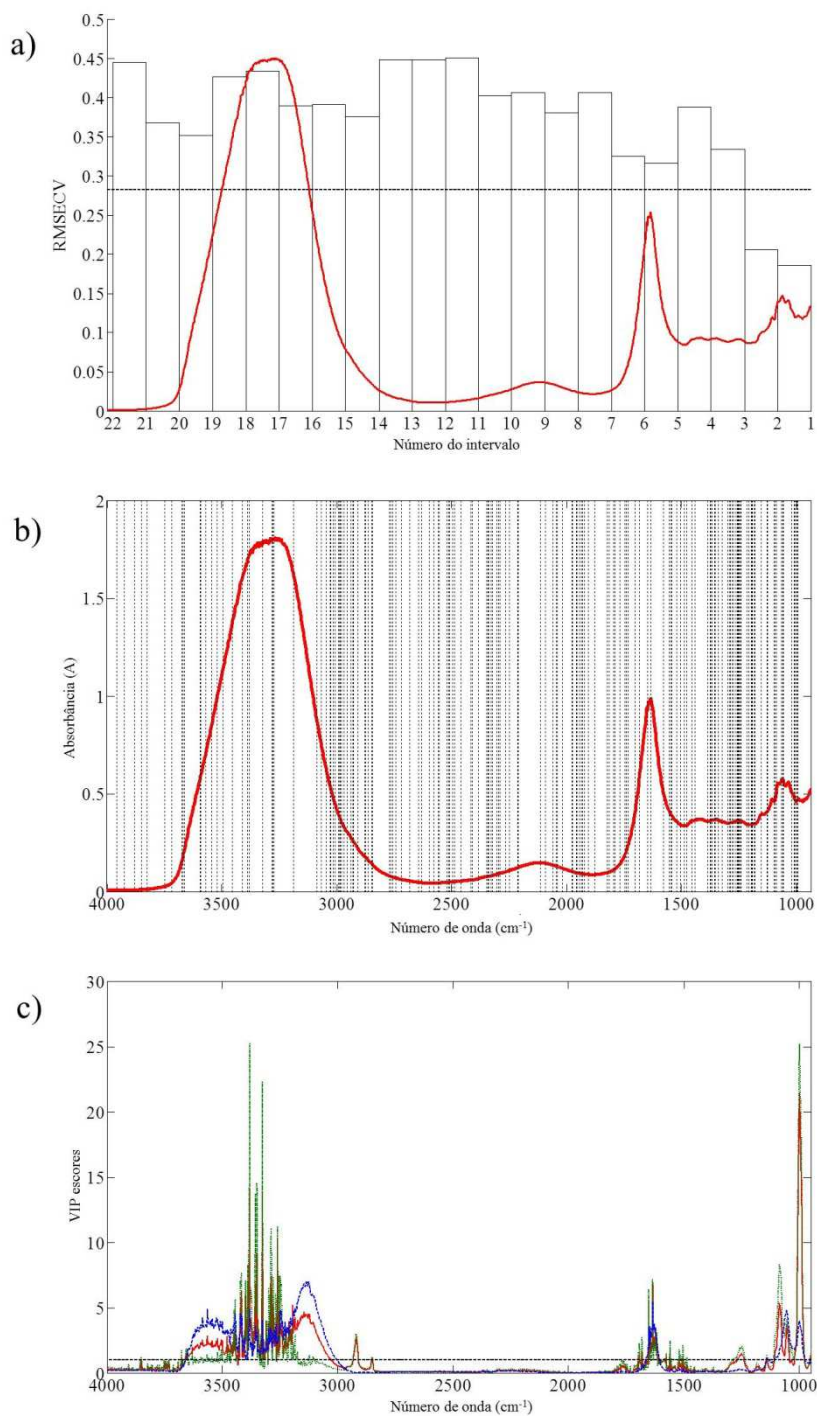
Com relação à predição dos dois conjuntos externos, contendo mistura dos dois adulterantes, em uma situação ideal, esperava-se que todas as amostras fossem atribuídas a ambas as classes adulteradas, CAS e APP, independente das porcentagens de adulteração. No modelo SIMCA, as amostras do conjunto externo 2, com concentrações mais altas de adulterantes (10 % de cada), foram, em sua maioria, classificadas na classe CAS (38 de 42), mas não na classe APP (apenas uma de 42). As amostras do conjunto externo 1, com a menor concentração de adulterantes (5 % de cada), não foram atribuídas a nenhuma classe.

No modelo PLS-DA, 5 de 15 amostras do conjunto externo 1 foram corretamente atribuídos às duas classes, CAS e APP, sendo um total de 11 amostras classificadas na classe CAS e 7 na APP. Apenas uma amostra do conjunto externo 2 foi corretamente atribuída às duas classes, enquanto todas as outras foram atribuídas somente à classe CAS.

III.4.4 PREDIÇÕES COM APLICAÇÃO DOS MÉTODOS DE SELEÇÃO DE VARIÁVEIS

Para melhorar os modelos anteriores, foi realizada uma otimização com os métodos de seleção de variáveis iPLS, GA e VIP scores, a fim de selecionar as variáveis/números de onda mais discriminantes entre as três classes. Os resultados das variáveis selecionadas por cada método estão ilustradas na **Figura III.4**.

Figura III.4 Gráfico das variáveis selecionadas por a) iPLS, com valores de RMSECV de cada intervalo, b) GA e c) VIP escores obtidos para as três classes



Legenda Figura III.4 a) linha tracejada horizontal indica o valor de RMSECV de referência obtido para o modelo construído com todas as variáveis espectrais; c): classe não adulterada (linha vermelha), classe adulterada com caju (linha verde) e classe adulterada com maçã (linha azul).

Para a seleção com iPLS, os espectros completos foram divididos em 20 subintervalos de 75 variáveis. A **Figura III.4a** mostra os valores de RMSECV obtidos para cada intervalo. A melhor capacidade preditiva foi observada para os intervalos 1 e 2, que correspondem a 150 variáveis/números de onda entre 1514 cm^{-1} e 937 cm^{-1} . Esses números de onda podem estar associados a sinais de açúcares e ácidos orgânicos presentes na região da impressão digital dos espectros ATR-FTIR dos néctares de frutas (Kelly & Downey, 2005; He, Rodriguez-Saona & Giusti, 2007; Bureau *et al.*, 2009).

Para a seleção com GA, foram aplicados os seguintes parâmetros: população de 64, largura da janela de 1, termos iniciais de 10, gerações máximas de 100, convergência de 50 %, taxa de mutação de 0,005 e cruzamento duplo. Parâmetros de validação cruzada foram definidos aleatoriamente com 20 divisões e 3 iterações. As 112 variáveis selecionadas estão distribuídas por todo o espectro (**Figura III.4b**).

Para a seleção com VIP scores, calculou-se a função quadrática dos pesos PLS para cada variável, considerando os modelos das três classes estudadas. Um total de 524 números de onda (**Figura III.4c**) apresentou VIP scores superiores a 1,0. Portanto, eles foram mantidos nos modelos otimizados.

Os modelos SIMCA e PLS-DA foram construídos em combinação com os métodos de seleção de variáveis. Os resultados obtidos para SIMCA estão mostrados na **Tabela III.2**. O iPLS melhorou significativamente as previsões dos conjuntos treinamento e teste em comparação com aqueles obtidos usando os espectros completos (**Tabela III.1**), fornecendo poucas amostras inconclusivas e sem erros nas atribuições. No entanto, para a previsão dos dois conjuntos de dados externos, nenhuma amostra foi corretamente atribuída às duas classes às quais eles pertenciam. Nenhuma melhoria significativa foi observada nos modelos construídos com as variáveis selecionadas pelo GA e VIP scores em relação aos modelos construídos com os espectros completos.

Tabela III.2 Resultados do modelo SIMCA combinado com os métodos de seleção de variáveis iPLS, GA e VIP escores

Método	Conjunto	Classe	Classificado como						
			UN	CAS	APP	NA	MA	IR	
iPLS	Treinamento	UN	25	0	0	3	0	3	
		CAS	0	26	0	2	0	2	
		APP	0	0	24	4	0	4	
	Teste	UN	13	0	0	1	0	1	
		CAS	0	14	0	0	0	0	
		APP	0	0	14	0	0	0	
	CAS + APP	1	0	0	0	15	0	15	
		2	0	0	0	42	0	42	
	GA	Treinamento	UN	26	28	0	0	26	26
			CAS	2	24	0	4	2	6
			APP	0	0	24	4	0	4
		Teste	UN	14	14	0	0	14	14
CAS			4	14	0	0	4	4	
APP			0	0	14	0	0	0	
CAS + APP		1	0	0	0	15	0	15	
		2	0	32	1	9	0	9	
VIP escores		Treinamento	UN	26	28	0	0	26	26
			CAS	7	21	0	7	7	14
			APP	0	0	22	6	0	6
		Teste	UN	14	14	0	0	14	14
	CAS		9	14	0	0	9	9	
	APP		0	0	14	0	0	0	
	CAS + APP	1	0	0	0	15	0	15	
		2	0	30	0	12	0	12	

Legenda: APP: classe adulterada com maçã; CAS: classe adulterada com caju; CAS + APP 1: adulterado com 5 % de maçã e 5 % de caju; CAS + APP 2: adulterado com 10 % de maçã e 10 % de caju; GA: algoritmo genético; IR: resultados inconclusivos; iPLS: regressão de mínimos quadrados parciais por intervalo; NA: não atribuído; MA: múltiplas atribuições; UN: classe não adulterada; VIP escores: variáveis importantes na projeção dos escores.

A **Tabela III.3** apresenta os resultados obtidos com PLS-DA combinado com os métodos de seleção de variáveis. Em geral, a classificação foi melhorada em relação ao modelo com os espectros completos (**Tabela III.1**). Todas as amostras nos conjuntos treinamento e teste foram corretamente classificadas com o iPLS. Os modelos construídos com GA e VIP escores, apresentaram resultados altamente satisfatórios, com apenas duas das 42 amostras não atribuídas a nenhuma classe, e sem atribuições erradas.

Tabela III.3 Resultados do modelo PLS-DA combinado com os métodos de seleção de variáveis iPLS, GA e VIP escores

Método	Conjunto	Classe	Classificado como						
			UN	CAS	APP	NA	MA	IR	
iPLS	Treinamento	UN	28	0	0	0	0	0	
		CAS	0	28	0	0	0	0	
		APP	0	0	28	0	0	0	
	Teste	UN	14	0	0	0	0	0	
		CAS	0	14	0	0	0	0	
		APP	0	0	14	0	0	0	
	CAS + APP	1	0	11	6	0	2	0	
		2	0	42	39	0	39	0	
	GA	Treinamento	UN	28	0	0	0	0	0
			CAS	0	28	0	0	0	0
APP			0	0	28	0	0	0	
Teste		UN	14	0	0	0	0	0	
		CAS	0	12	0	2	0	2	
		APP	0	0	14	0	0	0	
CAS + APP		1	0	14	4	1	4	1	
		2	0	42	2	0	2	0	
VIP escores		Treinamento	UN	28	0	0	0	0	0
			CAS	0	27	0	1	0	1
	APP		0	0	28	0	0	0	
	Teste	UN	14	0	0	0	0	0	
		CAS	0	13	0	1	0	1	
		APP	0	0	14	0	0	0	
	CAS + APP	1	3	11	6	0	5	0	
		2	0	42	1	0	1	0	

Legenda: APP: classe adulterada com maçã; CAS: classe adulterada com caju; CAS + APP 1: adulterado com 5 % de maçã e 5 % de caju; CAS + APP 2: adulterado com 10 % de maçã e 10 % de caju; GA: algoritmo genético; IR: resultados inconclusivos; iPLS: regressão de mínimos quadrados parciais por intervalo; NA: não atribuído; MA: múltiplas atribuições; UN: classe não adulterada; VIP escores: variáveis importantes na projeção dos escores.

Para o conjunto externo 2, apenas o iPLS forneceu resultados satisfatórios para a classificação de amostras em ambas as classes adulteradas (39 de 42). Quando o GA e VIP escores foram usados, as amostras foram principalmente atribuídas apenas à classe CAS. Para o conjunto externo 1, não foram obtidos melhores resultados com as variáveis selecionadas pelas três técnicas de seleção e os resultados foram similares aos obtidos com os espectros completos.

A principal diferença entre as estratégias de seleção de variáveis implementadas está no fato de que, embora o iPLS forneça uma seleção de blocos contínua, o GA e VIP escores selecionam variáveis discretas individuais. Assim, uma otimização mais exaustiva foi realizada aplicando a seleção de variáveis GA e VIP escores para a construção de modelos PLS-DA com os 150 números de onda selecionados pelo iPLS e estão dispostos na **Tabela III.4**. Os mesmos parâmetros descritos anteriormente para GA foram utilizados, exceto pelo número de iterações, que foi igual a 1. O número de variáveis selecionadas pela combinação do iPLS com GA e VIP escores foi de 16 e 84, respectivamente. Um resultado ligeiramente melhor foi obtido com o iPLS-GA, uma vez que todas as amostras do conjunto externo 2 foram corretamente atribuídas às classes CAS e APP e 6 de 15 amostras do conjunto externo 1 foram corretamente atribuídas às duas classes.

Tabela III.4 Resultados do modelo PLS-DA combinado com os métodos de seleção de variáveis iPLS-GA e iPLS-VIP

Método	Conjunto	Classe	Classificado como					
			UN	CAS	APP	NA	MA	IR
iPLS-GA	Treinamento	UN	28	0	0	0	0	0
		CAS	0	28	0	0	0	0
		APP	0	0	28	0	0	0
	Teste	UN	14	0	0	0	0	0
		CAS	0	14	0	0	0	0
		APP	0	0	14	0	0	0
	CAS + APP	1	0	11	10	0	6	0
		2	0	42	42	0	42	0
	iPLS-VIP	Treinamento	UN	28	0	0	0	0
CAS			0	28	0	0	0	0
APP			0	0	28	0	0	0
Teste		UN	14	0	0	0	0	0
		CAS	0	14	0	0	0	0
		APP	0	0	14	0	0	0
CAS + APP		1	0	9	4	3	1	3
		2	0	42	29	0	29	0

Legenda: APP: classe adulterada com maçã; CAS: classe adulterada com caju; CAS + APP 1: adulterado com 5 % de maçã e 5 % de caju; CAS + APP 2: adulterado com 10 % de maçã e 10 % de caju; GA: algoritmo genético; IR: resultados inconclusivos; iPLS: regressão de mínimos quadrados parciais por intervalo; NA: não atribuído; MA: múltiplas atribuições; UN: classe não adulterada; VIP: variáveis importantes na projeção dos escores.

Em geral, o PLS-DA apresentou um desempenho de classificação consistentemente melhor que o SIMCA. Assim, a **Tabela III.5** reporta uma comparação entre o modelo com espectro completo e os cinco modelos diferentes de seleção de variáveis com base nas figuras de mérito apropriadas. SEN e SEL foram estimadas como parâmetros globais, considerando todos os conjuntos de dados analisados. Uma SEL de 100 % foi observada para todos os casos, exceto

para a classe UN, quando prevista pelos modelos completo e VIP escores (SEL de 98 % e 96 %, respectivamente). Este resultado (SEL de 100 %) indica que nenhuma amostra foi atribuída a uma classe à qual não pertence. Todos os modelos da **Tabela III.5** forneceram uma SEN de 100 % para a classe UN e acima de 92 % para a classe CAS. No entanto, apenas modelos baseados na seleção por iPLS apresentaram uma SEN aceitável para APP. Os melhores valores de SEN foram 83 % para iPLS e 93 % para iPLS-GA. Estes resultados demonstram a dificuldade de prever amostras contendo ambos os adulterantes como pertencentes à classe APP. Em outras palavras, só foi possível detectar eficientemente a adulteração de maçã na presença simultânea de caju como adulterante empregando a seleção de variáveis por iPLS.

Tabela III.5 Sensibilidade (SEN) e seletividade (SEL) para os modelos PLS-DA com espectro completo e com os diferentes métodos de seleção de variáveis

Método	Classe					
	UN		CAS		APP	
	SEN	SEL	SEN	SEL	SEN	SEL
Completo	100	98	92	100	31	100
iPLS	100	100	94	100	83	100
GA	100	100	95	100	28	100
VIP escores	100	96	93	100	30	100
iPLS-GA	100	100	94	100	93	100
iPLS-VIP	100	100	92	100	66	100

Legenda: APP: classe adulterada com maçã; CAS: classe adulterada com caju; GA: algoritmo genético; iPLS: regressão de mínimos quadrados parciais por intervalo; iPLS-GA: regressão de mínimos quadrados parciais por intervalo e GA; iPLS-VIP: regressão de mínimos quadrados parciais por intervalo e VIP; UN: classe não adulterada; VIP escores: variáveis importantes na projeção dos escores.

III.5 CONCLUSÃO

O PLS-DA apresentou melhores resultados do que o SIMCA, uma vez que este forneceu um grande número de resultados inconclusivos. O modelo PLS-DA construído com o espectro completo foi capaz de classificar corretamente amostras dos conjuntos treinamento e teste, com bons parâmetros de desempenho. No entanto, este modelo não foi capaz de atribuir corretamente amostras adulteradas simultaneamente com caju e maçã nas suas respectivas classes. Esta tarefa só foi possível para misturas em um nível de concentração mais alto (conjunto externo 2) empregando métodos de seleção de variáveis.

Portanto, ao lidar com misturas, é interessante implementar diferentes estratégias que permitam melhorar os resultados em níveis mais baixos de concentração.

O melhor modelo foi obtido combinando os dois intervalos de onda mais preditivos selecionados pelo iPLS. Uma pequena melhora foi obtida quando o iPLS foi combinado com o algoritmo genético. Os números de onda mais preditivos estão localizados na região da impressão digital (entre 1200 e 950 cm^{-1}) e podem estar relacionados a ácidos orgânicos e açúcares. Assim, estes componentes nos néctares de frutas podem ser considerados como discriminantes entre as diferentes frutas utilizadas para a fabricação e adulteração deste tipo de bebida.

Do ponto de vista analítico, o aspecto mais interessante deste estudo é fornecer uma estratégia para construir modelos robustos de classificação multivariada capazes de detectar especificamente mais de um adulterante na presença de uma mistura deles. Este assunto é relativamente subestimado na literatura quimiométrica. Este estudo mostrou as limitações dos atuais métodos de classificação para concluir essa tarefa, caso não tenham sido especificamente projetados para esse fim. Além de alimentos e bebidas, essa estratégia pode ser aplicada na detecção de fraudes e adulterações em outros tipos de matrizes, como produtos farmacêuticos, combustíveis e amostras forenses.

CAPÍTULO IV. COMPARAÇÃO DE DIFERENTES MÉTODOS DE CLASSIFICAÇÃO MULTIVARIADA PARA DETECÇÃO DE ADULTERAÇÕES EM NÉCTARES DE UVA POR RESSONÂNCIA MAGNÉTICA NUCLEAR NO DOMÍNIO DE TEMPO (RMN-DT)

RESUMO

A adulteração nos néctares de uva por substituição da fruta principal por suco de maçã, suco de caju e uma mistura de maçã e caju foram avaliadas pela aplicação de ressonância magnética nuclear no domínio tempo (RMN-DT) associada a métodos de classificação multivariada. Duas abordagens diferentes foram investigadas: classe única (em que apenas as amostras não adulteradas foram modeladas) e multiclasses (três classes foram modeladas: classe não adulterada - UN, classes adulteradas com caju - CAS e classe adulterada com maçã - APP). Para a abordagem classe única, a modelagem independente e flexível por analogia de classe (SIMCA), regressão de mínimos quadrados parciais de uma classe (OCPLS) e SIMCA baseada em dados (DD-SIMCA) foram usados. Para a abordagem multiclasses, utilizou-se análise discriminante por mínimos quadrados parciais (PLS-DA) e SIMCA multiclasse. Os resultados obtidos demonstraram bom desempenho de todos os métodos de classe única com valores de eficiência superiores a 93 %. Para a abordagem multiclasses, a predição das amostras contendo apenas um tipo de adulterante apresentou eficiência superior a 90 % e 97 %, usando SIMCA e PLS-DA, respectivamente. A predição das amostras contendo mistura dos dois adulterantes foi satisfatória para a classe CAS, mas não para a classe APP, utilizando PLS-DA. No entanto, o modelo SIMCA multiclasse não forneceu previsões satisfatórias para nenhuma dessas duas classes.

Palavras-chave: RMN de baixo campo. Modelos de classe única. Modelos multiclasses. Néctar de fruta. Adulteração em alimentos.

IV.1 PRODUÇÃO

Miaw, C. S. W.; Santos, P. M.; Silva, A. R. C. S.; Gozzi, A.; Guimarães, N. C. C.; Callao, M. P.; Ruisánchez, I.; Sena, M. M.; Souza, S. V. C. Comparison of different multivariate classification methods for the detection of adulterations in grape nectars by using low-field nuclear magnetic resonance. Manuscrito submetido à revista **Food Control**.

IV.2 INTRODUÇÃO

A ressonância magnética nuclear no domínio do tempo (RMN-DT) é uma técnica analítica particularmente atrativa para triagem em análise de alimentos. Ela é primariamente baseada nas medidas de relaxação e difusão do sinal provindo da amostra. A técnica é chamada de RMN-DT, porque os sinais são obtidos no domínio do tempo, monitorando os tempos de relaxação longitudinal (T_1) e transversal (T_2) (Colnago & Andrade, 2016). A maioria das análises é realizada com a sequência de pulsos Carr-Purcell-Meiboom-Gill (CPMG), produzindo um decaimento exponencial com a constante de tempo T_2 , que também é chamada de tempo de relaxamento *spin-spin* (Shao & Li, 2012; Kirtil *et al.*, 2017). Os espectrômetros RMN-DT utilizam imãs livres de fluidos criogênicos com equipamento de bancada relativamente barato, rápido e fácil de usar (Kirtil *et al.*, 2017). Além disso, eles admitem medidas diretas, através de embalagens de vidro ou de plástico, permitindo que, no futuro, os consumidores possam checar a qualidade do produto na embalagem antes da compra (Santos & Colnago, 2018).

Nas últimas duas décadas, um aumento acentuado nas aplicações de RMN-DT foi observado, com foco em análises relacionadas à distribuição e estado da água ou óleo/gordura. Sinais obtidos nesses experimentos produzem conjuntos de dados grandes e colineares que não podem ser tratados por métodos univariados tradicionais, exigindo, assim, métodos quimiométricos. Esses métodos podem modelar tendências multivariadas associadas às diferenças nos teores de água e/ou gordura em alimentos. Eles extraem mais informações de decaimentos CPMG do que os valores discretos de T_2 obtidos ajustando os dados com uma ou mais curvas exponenciais. Eles também evitam a necessidade de pré-secagem da amostra ou a utilização de gradiente de campo pulsado e medições em um equilíbrio dinâmico em estado estacionário (Todt *et al.*, 2006).

Foram encontrados trabalhos aplicando RMN-DT combinado a métodos multivariados qualitativos e quantitativos na literatura recente analisando milho doce (Shao & Li, 2013), ostras (Davenel *et al.*, 2009), ameixas (Pereira *et al.*, 2013a),

carne bovina (Pereira *et al.*, 2013b), leite (Santos, Pereira-Filho & Colnago, 2016) e geléias (Santos & Colnago 2018). Em relação à análise de bebidas de frutas, a espectroscopia RMN-DT aliada à quimiometria foi utilizada para prever e classificar laranjas a partir de seus parâmetros internos de qualidade e avaliar os atributos sensoriais dos sucos de laranja (Flores *et al.*, 2016). Não foram encontrados na literatura trabalhos utilizando RMN-DT para avaliar a autenticação/adulteração em outras bebidas à base de frutas.

O objetivo deste estudo foi aplicar espectroscopia RMN-DT e métodos de classificação multivariada para avaliar a autenticidade de néctares de uva adulterados com suco de caju, suco de maçã e a mistura dos dois adulterantes. A adulteração com misturas de dois ou mais adulterantes pode ocorrer em algumas situações reais, como é o caso da adulteração de néctar. Portanto, é importante verificar se o método desenvolvido para identificar a presença de um adulterante é capaz de classificar corretamente as amostras adulteradas com mais de um.

Uma amostra analisada pode ser atribuída a uma, mais de uma ou a nenhuma classe (Gondim *et al.*, 2017). Em tese, amostras adulteradas com mais de um adulterante devem ser atribuídas a todas as classes estabelecidas para os respectivos adulterantes. As abordagens de classe única e multiclases foram aplicadas. Na abordagem de classe única, apenas as amostras não adulteradas (UN) foram consideradas como a classe autêntica (classe alvo). Com este propósito, foram construídos modelos usando SIMCA, mínimos quadrados parciais de uma classe (OCPLS) e SIMCA orientado aos dados (DD-SIMCA).

Na abordagem multiclases, além da classe não adulterada, classes adulteradas específicas foram definidas. As classes UN, CAS e APP foram modeladas e aplicaram-se os métodos de classificação SIMCA e PLS-DA.

IV.3 MATERIAL E MÉTODOS

IV.3.1 MÉTODOS

IV.3.1.1 Análise de dados

Para os modelos SIMCA e DD-SIMCA, as curvas de relaxação de T_2 foram primeiramente normalizadas (normalizadas pela variável que apresenta máxima intensidade) e centradas na média, enquanto que nenhum destes pré-processamentos foi aplicado no modelo OCPLS. As análises por PCA, SIMCA e PLS-DA foram realizadas utilizando o *software* MATLAB, versão 7.10.0.499 (Natick, MA, EUA) e PLS Toolbox, versão 6.7.1 (Eigenvector Research Inc., Wenatchee, WA, EUA). A rotina do OCPLS foi fornecida por Lu Xu e do DD-SIMCA por Alexey L. Pomerantsev.

IV.3.1.2 Construção dos modelos de classificação

As amostras de cada classe usadas para construir os modelos (UN, CAS, APP) foram divididas em conjuntos de treinamento e teste usando o algoritmo Kennard-Stone (28 para o treinamento e 14 para o conjunto de testes). Para o modelo de classe única, o conjunto teste de amostras não adulteradas e todas as amostras adulteradas CAS e APP (conjuntos treinamento e teste) e as misturas (CAS + APP) foram usadas para previsão. Para a abordagem multiclases, os conjuntos teste das três classes e as misturas foram usados para previsão.

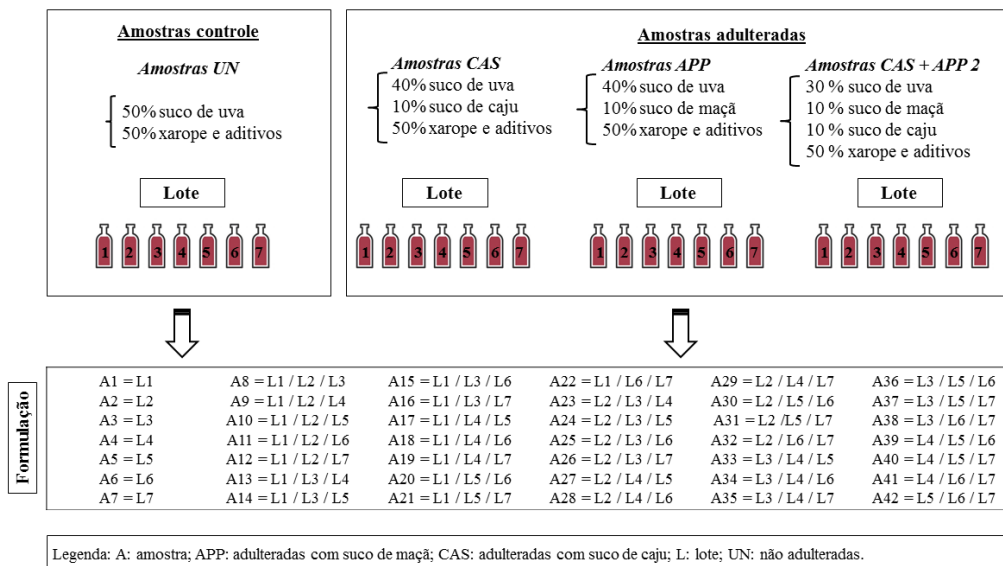
IV.3.2 DELINEAMENTO

IV.3.2.1 Formulação dos néctares

O delineamento da formulação de amostras não adulteradas (UN) e adulteradas está ilustrada na **Figura IV.1**. Foram obtidos três tipos de amostras adulteradas: néctares adulterados com caju (CAS), adulterados com maçã (APP) e adulterados com caju e maçã (CAS + APP). As amostras foram armazenadas em tubos Falcon de 50 mL no freezer (-18 a -24 °C).

No **Capítulo III**, foi necessário o uso de seleção de variáveis para aumentar a capacidade preditiva dos modelos construídos a partir da técnica espectroscópica ATR-FTIR, a fim de prever as amostras contendo as misturas dos adulterantes caju e maçã. As amostras adulteradas com as misturas no nível mais baixo de adulterante (5 % de cada) tiveram poucas atribuições corretas em ambas às classes (CAS e APP). Dessa forma, optou-se por prever somente as amostras contendo os níveis mais altos de adulterantes (10 % de cada) neste estudo, uma vez que não seria possível aplicar métodos de seleção de variáveis a partir das curvas de relaxação T_2 obtidas com a técnica RMN-DT.

Figura IV.1 Delineamento experimental da formulação dos néctares de uva



Legenda: A: amostra; APP: classe adulterada com suco de maçã; CAS: classe adulterada com suco de caju; CAS + APP 2: amostras adulteradas com caju e maçã com 10 % de cada adulterante; L: lote; UN: classe não adulterada

IV.3.2.2 Parâmetros de desempenho

Para determinar a qualidade dos métodos de classificação, os parâmetros de desempenho sensibilidade (SEN), seletividade (SEL), resultados inconclusivos (IR) e eficiência (EFF) foram calculados.

Na abordagem de classe única, foram calculados os parâmetros SEN, SEL e EFF, enquanto que na abordagem multiclasses, além dos três parâmetros mencionados, também foi calculado o número de IR.

IV.4 RESULTADOS E DISCUSSÃO

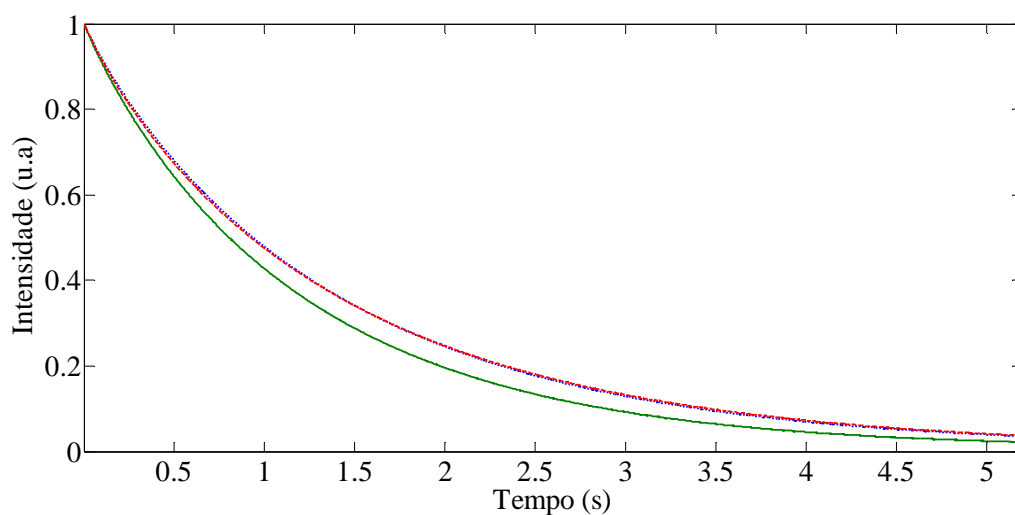
IV.4.1 RESULTADOS DE RELAXAÇÃO DE T_2

As médias das curvas de relaxação de T_2 normalizadas das amostras das classes UN, CAS e APP estão representadas na **Figura IV.2a**. Como pode ser observado, não houve diferença significativa no tempo T_2 entre as amostras das classes UN e APP. Por outro lado, as amostras da classe CAS decaíram mais rapidamente, apresentando um T_2 mais curto quando comparado com as amostras de UN e APP.

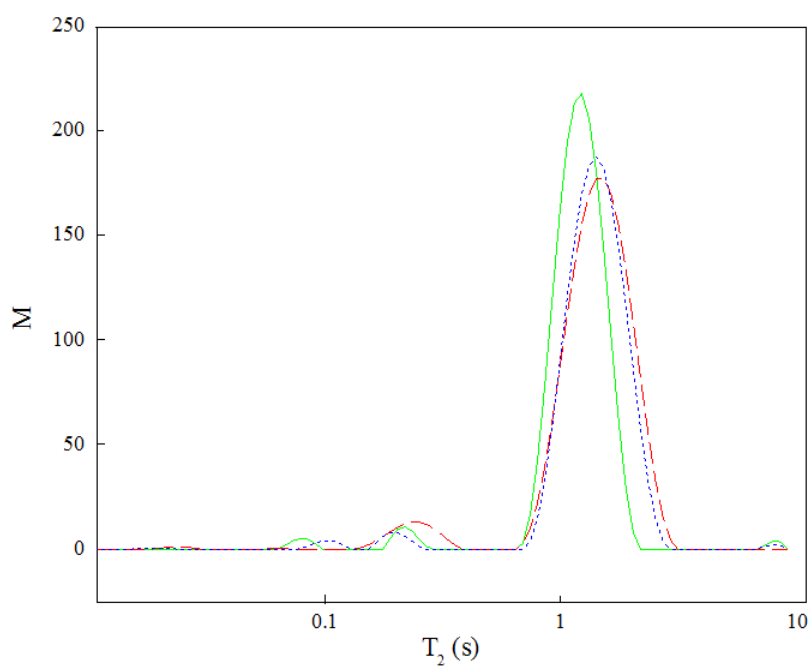
Para verificar o número de componentes nas amostras, uma transformada inversa de Laplace foi aplicada nos decaimentos CPMG. Os tempos de relaxação T_2 distribuídos (espectro de relaxação) para as amostras das classes UN, CAS e APP estão ilustrados na **Figura IV.2b**. Para amostras de CAS (linha verde), os tempos de relaxação T_2 distribuídos são caracterizados por três populações de relaxação; dois componentes menores em torno de 0,08 s e 0,21 s, respectivamente, e o maior componente em torno de 1,23 s. As amostras de UN (linha azul) também são caracterizadas por três populações de relaxação em torno de 0,10, 0,20 e 1,49 s, enquanto amostras de APP (linha vermelha) mostraram apenas duas populações de relaxação centradas em 0,24 s e 1,48 s. Em todas as amostras, a maior população (com o T_2 maior) representa mais da metade da população total de prótons nas amostras de néctares e está associada à água mais móvel.

Figura IV.2 a) Curvas de relaxação obtidas com a normalização das médias CPMG para as amostras não adulteradas (linha azul pontilhada), adulteradas com caju (linha verde sólida) e adulteradas com maçã (linha vermelha tracejada). b) Distribuição dos tempos de relaxação T_2 estimados pela transformada inversa de Laplace das curvas CPMG obtidas para amostra não adulterada (linha azul pontilhada), adulteradas com caju (linha verde sólida) e adulteradas com maçã (linha vermelha tracejada)

a)



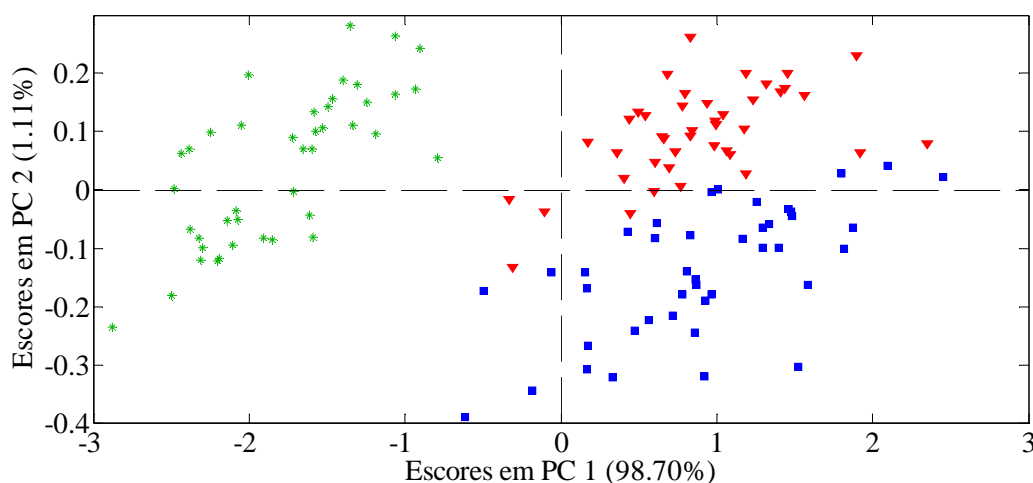
b)



IV.4.2 MODELO PCA

A **Figura IV.3** mostra a visão geral da variação dos dados de relaxação de RMN-DT (amostras UN, CAS e APP) através de uma PCA realizada nos dados CPMG normalizados e centrados na média. As duas primeiras PC explicaram quase toda a variância (99,8 %). PC1 explicou 98,70 % da variância total, claramente discriminando amostras CAS (valores de escores negativos) das amostras UN e APP (valores de escores positivos). Essa discriminação de classes é a mais importante no conjunto de dados corroborando a observação direta das curvas de relaxação T_2 (Figura IV.2 a). As amostras UN e APP foram separadas apenas pela projeção de seus escores em PC2, que explicou 1,11 % da variância total.

Figura IV.3 Análise de componentes principais dos dados de relaxação CPMG das amostras de néctares de uva não adulteradas (▼), adulteradas com caju (★) e adulteradas com maçã (■)



IV.4.3 MÉTODOS DE CLASSE ÚNICA

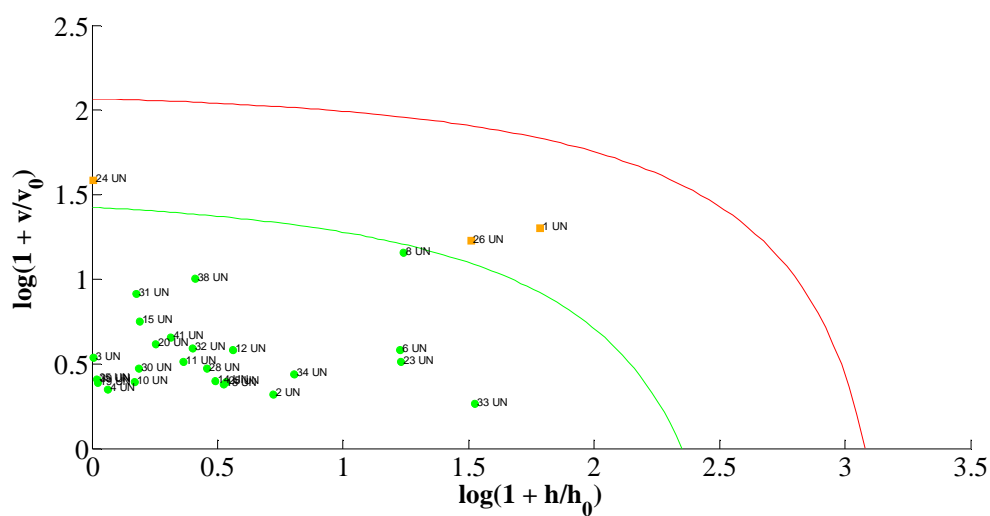
Na abordagem de classe única, apenas a classe não adulterada foi modelada. O melhor modelo SIMCA foi construído com base na PCA das amostras não

adulteradas. Uma PC foi selecionada por validação cruzada explicando 98,1 % da variância das amostras independentes.

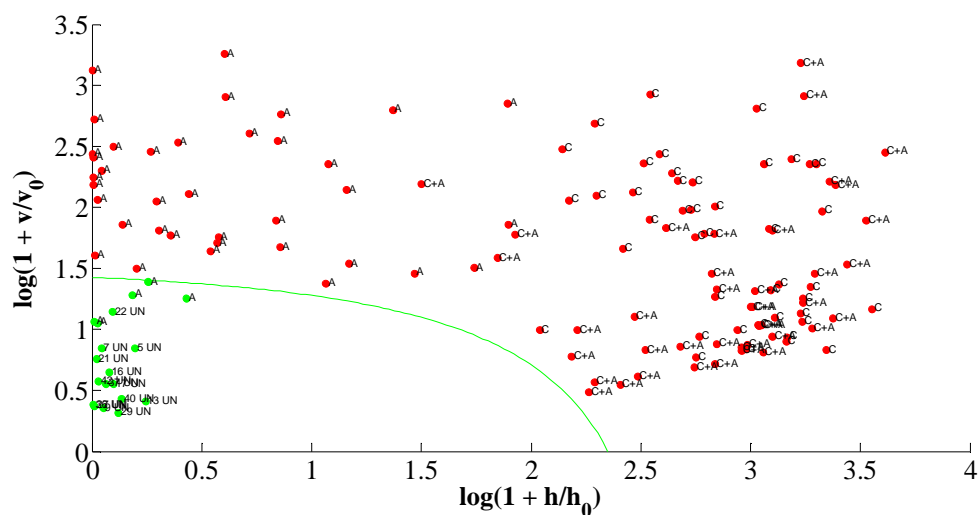
Para o modelo DD-SIMCA, apenas uma PC foi selecionada para modelar as amostras UN, o nível de significância para a detecção de *outlier* foi fixado em 0,01, a área de aceitação foi baseada na distribuição qui-quadrado e o erro tipo I (α) fixado em 0,05. A **Figura IV.4a** mostra o gráfico de aceitação para o conjunto de treinamento UN e a **Figura IV.4b** mostra o gráfico de aceitação para o conjunto de teste UN, para amostras CAS, amostras APP e amostras CAS+APP. A linha verde representa a fronteira/limite entre as amostras que realmente pertencem à classe UN e as amostras que não pertencem à classe, enquanto a linha vermelha representa o limite da área de aceitação.

Figura IV.4 Gráficos de aceitação obtidos com DD-SIMCA para os conjuntos treinamento (a) e teste (b). h_i e v_i são, respectivamente, os valores da distância da pontuação e da distância ortogonal para a amostra $i = 1, \dots, I$

a)



b)

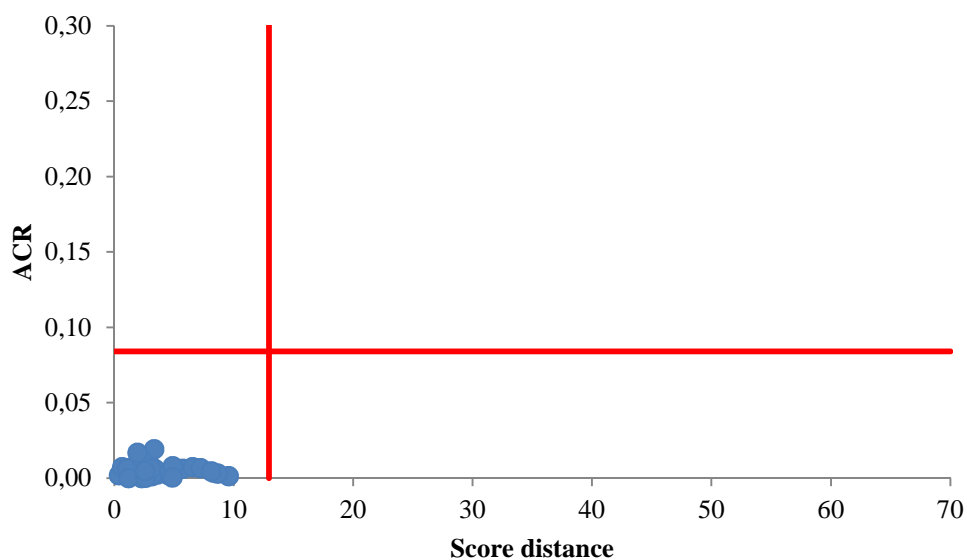


No modelo OCPLS, uma validação cruzada foi aplicada no conjunto treinamento e, como resultado, um vetor contendo desvios padrões do modelo residual com diferentes números de VLs foi obtido. O número ideal de VLs selecionado (3 VLs) foi aquele com o menor desvio padrão dos resíduos absolutos centralizados do modelo (*ACR – absolute centered residual*).

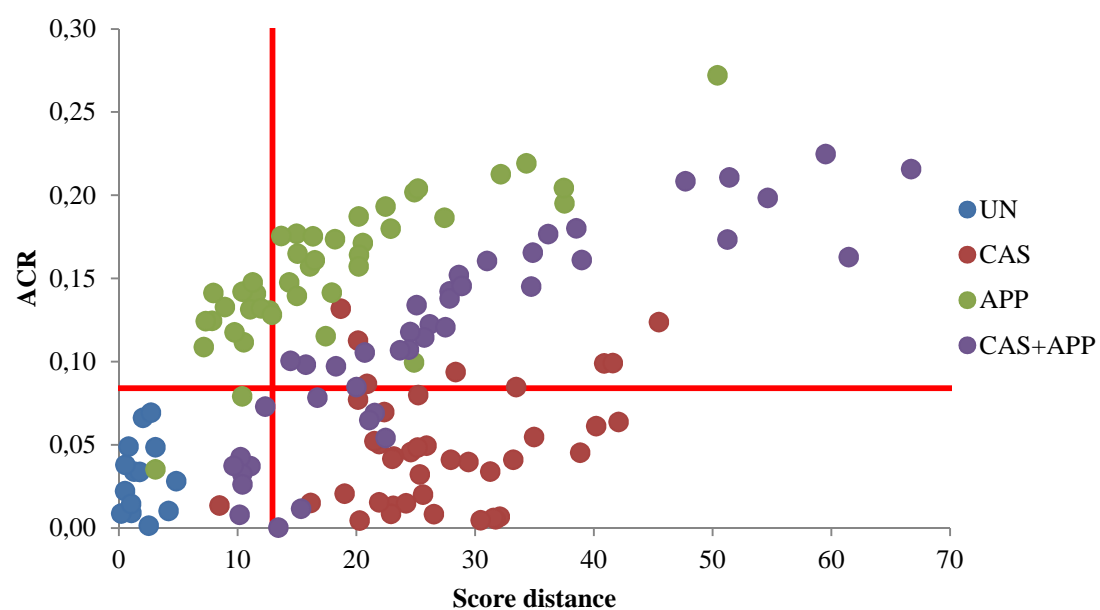
A **Figura IV.5a** mostra os limites (linha contínua vermelha) para a classe UN no conjunto treinamento. A **Figura IV.5b** mostra a previsão das amostras não adulteradas do conjunto teste e de todas as amostras adulteradas. Como pode ser observado, todas as amostras não adulteradas do conjunto de teste estão dentro dos limites do modelo, no quadrante inferior esquerdo. A maioria das amostras adulteradas está fora dos limites do modelo.

Figura IV.5 Gráficos obtidos pelo modelo OCPLS das distâncias do scores (SD) versus resíduos absolutos centralizados (ACR) das amostras do conjunto treinamento (a) e teste (b)

a)



b)



Para os três modelos, a SEN foi igual a 100 % no conjunto de treinamento. A **Tabela IV.1** apresenta os parâmetros de desempenho estimados para o conjunto teste das amostras UN e para todas as amostras adulteradas. Valores próximos de EFF foram obtidos para os modelos SIMCA e DD-SIMCA, de 97,1 e 96,4 %, respectivamente, enquanto que um valor menor, 92,9 %, foi obtido para OCPLS. A partir destes resultados, pode se concluir que todos os modelos foram capazes de reconhecer as amostras UN (100 % de SEN), apesar dos valores de SEL serem menores que 100 %. Tanto SIMCA como DD-SIMCA apresentaram valores de SEL baixos apenas para classe APP (entre 88,1 e 90,5 %), enquanto que o OCPLS apresentou valores baixos (entre 83,3 e 97,6 %) independentemente dos adulterantes estarem presentes de forma isolada ou como mistura.

Tabela IV.1 Parâmetros de desempenho estimados a partir dos conjuntos teste das amostras não adulterados e todas as amostras adulteradas para os modelos de classe única SIMCA, OCPLS e DD-SIMCA

Método	Amostras	SEN (%)	SEL (%)	EFF (%)
SIMCA	UN	100,0	-	97,1
	CAS	-	100,0	
	APP	-	90,5	
	CAS + APP	-	100,0	
OCPLS	UN	100,0	-	92,9
	CAS	-	97,6	
	APP	-	95,2	
	CAS + APP	-	83,3	
DD-SIMCA	UN	100,0	-	96,4
	CAS	-	100,0	
	APP	-	88,1	
	CAS + APP	-	100,0	

Legenda: APP: classe adulterada com maçã; CAS: classe adulterada com caju; CAS+APP: amostras adulteradas com caju e maçã; UN: classe não adulterada.

IV.4.4 MODELOS MULTICLASSES

Os modelos multiclases SIMCA e PLS-DA foram construídos com o conjunto de treinamento das três classes (UN, CAS e APP). O conjunto de teste e as amostras das mistura (CAS+APP) foram utilizados, de forma independente, para validação dos métodos.

No modelo SIMCA, cada classe foi modelada independentemente e o número ideal de PCs escolhido com base no menor valor de RMSECV. Uma PC para cada classe foi selecionada, representando 98,07, 96,46 e 98,01 % da variância total, para as classes UN, CAS e APP, respectivamente.

O modelo PLS-DA também foi construído com as três classes. Utilizou-se a validação cruzada por venezianas e o número de VLs foi escolhido com base nos menores erros de classificação de validação cruzada. Com 3 LVs e 99,84 % de variância no bloco **X** e 55,87 % no bloco **Y**, os valores de *threshold* obtidos foram 0,22 para a classe UN, 0,15 para a classe CAS e 0,12 para a classe APP.

A **Tabela IV.2** exibe os resultados dos dois métodos na forma de uma matriz de confusão. No geral, o PLS-DA apresentou melhores resultados do que o SIMCA.

Em relação ao método PLS-DA, na previsão do conjunto treinamento apenas uma amostra foi incorretamente atribuída (uma amostra UN foi classificada como APP). Além disso, somente duas atribuições inconclusivas das amostras UN (uma não atribuída e uma atribuição múltipla) foram obtidas. Todas as amostras adulteradas do conjunto de treinamento e teste foram corretamente classificadas.

Em relação ao método SIMCA, duas amostras UN foram atribuídas erroneamente à classe APP e atribuições inconclusivas (não atribuições e múltiplas atribuições) foram obtidas nas três classes modeladas, tanto no conjunto de treinamento como no de teste.

Tabela IV.2 Predições multiclases de SIMCA e PLS-DA das amostras da classe não adulterada (UN), adulterada com caju (CAS), adulterada com maçã (APP) para conjuntos de treinamento e teste e das amostras adulteradas com caju e maçã (CAS+APP)

Método	Conjunto	Classe	n	Classificado como					
				UN	CAS	APP	NA	MA	IR (%)
SIMCA	Treinamento	UN	28	25	0	5	1	3	14,3
		CAS	27*	0	25	0	2	0	7,4
		APP	28	3	0	25	3	3	21,4
	Teste	UN	14	14	0	2	0	2	7,1
		CAS	14	0	13	0	1	0	3,6
		APP	14	1	0	14	0	1	3,6
		CAS+APP	42	0	15	0	27	0	-
PLS-DA	Treinamento	UN	28	27	0	1	1	1	7,1
		CAS	27*	0	27	0	0	0	0,0
		APP	28	0	0	28	0	0	0,0
	Teste	UN	14	12	0	0	2	0	14,3
		CAS	14	0	14	0	0	0	0,0
		APP	14	0	0	14	0	0	0,0
		CAS+APP	42	0	41	18	0	17	-

Legenda: APP: classe adulterada com maçã; CAS: classe adulterada com caju; CAS+APP: amostras adulteradas com caju e maçã; n: número de amostras; NA: não atribuído; MA: atribuições múltiplas; IR: resultados inconclusivos; PLS-DA: análise discriminante por mínimos quadrados parciais; SIMCA: modelagem independente e flexível por analogia de classe; UN: classe não adulterada. *Uma amostra foi eliminada desta classe, pois foi considerada um *outlier*.

A classificação das amostras contendo a mistura dos adulterantes caju e maçã (CAS + APP 2) no modelo PLS-DA indicou que aproximadamente metade delas foram adequadamente atribuídas a ambas as classes (CAS e APP). O restante foi atribuído a apenas uma classe, principalmente à classe CAS. Para esse conjunto de dados, o SIMCA apresentou resultados piores. Nenhuma amostra foi classificada em ambas as classes simultaneamente, e apenas 15 amostras foram corretamente classificadas na classe CAS. Nenhuma amostra foi atribuída à classe APP.

A **Tabela IV.3** apresenta os parâmetros de desempenho para o conjunto de teste e o conjunto das misturas. Por uma questão de clareza, os parâmetros de desempenho foram calculados considerando as amostras contendo apenas um adulterante (CAS e APP) e, adicionalmente, sem diferenciar entre contendo um e mais de um adulterante (CAS, APP e CAS + APP), uma vez que o comportamento das amostras contendo misturas é relativamente diferente.

Tabela IV.3 Parâmetros de desempenho sensibilidade (SEN), seletividade (SEL), taxa de inconclusivos (IR) e eficiência (EFF) para modelos multiclasse SIMCA e PLS-DA

Método	AD	Classe											
		UN				CAS				APP			
		SEN	SEL	IR	EFF	SEN	SEL	IR	EFF	SEN	SEL	IR	EFF
SIMCA	CAS e APP	100	96	14	98	93	100	7	98	100	93	7	98
	CAS, APP e CAS+APP	100	99	14	98	50	100	50	67	25	93	50	49
	CAS e APP	86	100	14	95	100	100	0	100	100	100	0	100
PLS-DA	CAS, APP e CAS+APP	86	100	14	98	98	100	0	99	57	100	0	71

Legenda: AD: conteúdo adulterante; APP: classe adulterada com maçã; CAS: classe adulterada com caju; CAS+APP: amostras adulteradas com caju e maçã PLS-DA: análise discriminante por mínimos quadrados parciais; SIMCA: modelagem independente e flexível por analogia de classe; UN: classe não adulterada.

Exclusivamente para a previsão da classe UN, o SIMCA multiclases superou o PLS-DA e os resultados foram comparáveis àqueles obtidos com a estratégia de classe única. No entanto, essa informação complementar da classe foi obtida na forma de taxa de inconclusivos (IR). Uma IR de 14 % foi estimada para esta classe em ambas as estratégias, tornando necessária uma análise confirmatória subsequente.

Para as duas classes adulteradas, quando amostras contendo apenas um adulterante foram consideradas, todos os parâmetros de desempenho apresentaram excelentes valores (SEN e SEL próximas a 100 % e 0 % de IR) para o PLS-DA e com valores um pouco inferiores (SEN e SEL maiores que 93 %, e 7 % de IR) para o SIMCA. Porém, quando adicionalmente a estas amostras as misturas (CAS+APP) também foram consideradas, o comportamento foi diferente. Os parâmetros de desempenho para classe CAS foram altamente satisfatórios para o PLS-DA, mas apresentaram valores significativamente menores para o SIMCA (SEN e IR de 50 % e 67 % de EFF). Para a classe APP, melhores resultados foram obtidos com PLS-DA, no entanto o nível de atribuição das amostras CAS+APP não foi considerado aceitável (SEN de 57 %). Esta tendência foi muito pior para o modelo SIMCA.

Embora os valores dos parâmetros de desempenho não tenham sido considerados satisfatórios para a predição das amostras contendo a mistura de adulterantes, é importante reforçar que em nenhum caso amostras adulteradas foram classificadas como não adulteradas. Em outras palavras, nenhum erro falso-negativo foi encontrado. Isso é importante no contexto dos métodos de triagem para detecção de fraude alimentar, uma vez que essas amostras podem ser subsequentemente avaliadas por um método confirmatório mais trabalhoso.

Estes resultados foram comparados com os resultados obtidos no **Capítulo II** utilizando a espectroscopia ATR-FTIR. Em relação à estratégia multiclases (sem considerar as amostras contendo as misturas de adulterantes), os modelos SIMCA utilizando RMN-DT apresentaram eficiência variando de 93 % a 98 % e foram considerados superiores à técnica ATR-FTIR, com resultados variando de 64 % a

92 %. Resultados superiores também foram obtidos com PLS-DA, apresentando EFF variando de 95 % a 100 % para RMN-DT e de 83 % a 100 % para espectroscopia ATR-FTIR.

Com base nesses resultados, pode-se afirmar que a técnica RMN-DT é apropriada para detectar amostras contendo um adulterante de cada vez, mas para misturas de dois adulterantes, pesquisas adicionais devem ser realizadas.

A estratégia de fusão de dados combinando as técnicas ATR-FTIR e RMN-DT foi testada, porém, não foi considerada eficiente, uma vez que apresentou resultados inferiores ao modelo ATR-FTIR aplicado de forma isolada.

IV.5 CONCLUSÃO

A espectroscopia RMN-DT aliada a métodos de classificação multivariada foi utilizada para detectar adulterações em néctares de uva. Duas estratégias de classificação foram aplicadas e comparadas de acordo com seus parâmetros de desempenho e nível de informação que podem fornecer.

Para estabelecer os modelos de classe única, OCPLS, DD-SIMCA e SIMCA foram aplicados. De forma global, EFF maiores que 93 % foram obtidas utilizando SIMCA e DD-SIMCA, com desempenhos superiores ao OCPLS. Para estabelecer os modelos multiclassés, SIMCA e PLS-DA foram aplicados. A predição das amostras contendo apenas um adulterante apresentou resultados satisfatórios para as três classes com valores de EFF maiores que 93 %. A predição das amostras CAS+APP evidenciou nenhuma atribuição na classe APP utilizando SIMCA e resultados um pouco melhores com PLS-DA. Este fato demonstrou que a estratégia multiclassés não foi efetiva na predição das amostras contendo a mistura dos dois adulterantes. Apesar de que em nenhum caso amostras adulteradas foram classificadas como adulteradas.

Como conclusão deste estudo, as duas estratégias foram eficientes para autenticar néctares de uva. Quando o adulterante é conhecido, a estratégia multiclassés pode fornecer informações adicionais, no entanto também é necessário obter amostras representativas dos tipos de adulteração em estudo. Como observado neste estudo, nem sempre amostras contendo mais de um tipo de adulterante serão classificadas nas classes modeladas contendo estes adulterantes.

5 CONSIDERAÇÕES FINAIS

A falta de métodos de análise consolidados e de uma fiscalização mais efetiva podem tornar os néctares de frutas susceptíveis a fraudes com objetivo de ganho econômico por parte de indústrias fraudulentas. Néctares de uva e laranja, que são os mais consumidos no Brasil e no mundo, possuem apenas a quantidade mínima de polpa requerida como requisito do PIQ, sendo que as análises de outros parâmetros importantes como acidez total, açúcares totais e sólidos solúveis não são exigidas. Na análise dos néctares de caju, maçã, maracujá e pêssego foi possível constatar que algumas indústrias atendem aos requisitos do PIQ, porém próximos aos limites mínimos exigidos, o que pode ser interpretado como uma tentativa de redução dos custos. Além disso, algumas marcas apresentaram altos valores de açúcares totais, sugerindo que o conteúdo de fruta possa estar sendo reduzido e substituído por xarope de açúcar ou por outra fruta com teor de sólidos solúveis maior do que a fruta principal.

Nos **Capítulos I a III** foi utilizada a espectroscopia ATR-FTIR tanto em modelos de calibração (**Capítulo I**) como de classificação multivariada (**Capítulos II e III**), tanto de forma a quantificar o teor de fruta principal nos néctares, como também na detecção e classificação dos possíveis adulterantes (maçã e caju) presentes.

No **Capítulo I**, os melhores modelos para quantificar a concentração da fruta principal foram fornecidos pelo iPLS-OPS. Os métodos foram validados demonstrando adequação para o propósito de uso e podem ser aplicados por órgãos fiscalizadores a fim de avaliar o percentual de polpa contido nestes produtos. As variáveis mais preditivas estavam concentradas na região de *fingerprinting* e foram associadas aos açúcares e a absorções de ácidos orgânicos, sugerindo que o conteúdo desses componentes nos néctares é importante para discriminar entre as frutas principais e os adulterantes.

No **Capítulo II**, o método de classificação multivariada PLS-DA apresentou os melhores resultados (altas sensibilidade e seletividade) tanto na estratégia de classe

única como na multiclases para a detecção dos adulterantes maçã e caju, de forma isolada.

No **Capítulo III**, apenas com o uso dos métodos de seleção de variáveis iPLS e iPLS-GA foi possível atribuir as amostras contendo a mistura dos adulterantes caju e maçã (e na concentração mais alta) a ambas as classes (CAS e APP). Novamente, os números de onda mais preditivos estavam localizados na região entre 1200 e 950 cm^{-1} (*fingerprinting*).

No **Capítulo IV**, foi utilizada a técnica de RMN-DT e classificação multivariada. Todos os modelos de classe única (OCPLS, DD-SIMCA e SIMCA) e multiclases (PLS-DA e SIMCA) apresentaram resultados satisfatórios (EFF > 93 %) para classificar corretamente as amostras UN, CAS e APP. Os modelos multiclases também classificaram corretamente as amostras UN. No entanto, nos modelos multiclases, em que as amostras contendo mistura dos dois adulterantes (CAS+APP) deveriam ter sido classificadas em ambas as classes, isso não ocorreu. Assim, a estratégia de classe única foi considerada mais eficiente, classificando amostras não adulteradas como pertencentes à classe alvo (classe não adulterada) e as adulteradas como não pertencentes à classe alvo.

O uso das espectroscopias ATR-FTIR e RMN-DT combinadas a diferentes métodos quimiométricos indicou sua potencialidade na análise de fraudes em néctares de frutas. Além disso, a combinação com métodos de seleção de variáveis foi importante recurso para melhorar a interpretação e robustez, tanto dos modelos de calibração como de classificação multivariada. A validação dos métodos indicou adequação aos propósitos de uso, estimando as figuras de mérito específicas para os modelos quantitativos e para os modelos qualitativos. Assim, pode-se sugerir a utilização destes métodos por órgãos governamentais ou mesmo por laboratório independentes no controle de qualidade, a fim de garantir a autenticidade dos néctares de fruta.

REFERÊNCIAS BIBLIOGRÁFICAS

ABAD-GARCÍA, B. *et al.* Polyphenolic contents in Citrus fruit juices: authenticity assessment. **European Food Research and Technology**, v. 238, p. 803-818, 2014.

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA - ANVISA. **Resolução da Diretoria Colegiada nº 8, de 06 março de 2013**: Dispõe sobre a aprovação de uso de aditivos alimentares para produtos de frutas e de vegetais e geléia de mocotó. 2013.

ALAMAR, P. D. *et al.* Quality evaluation of frozen guava and yellow passion fruit pulps by NIR spectroscopy and chemometrics. **Food Research International**, v. 85, p. 209-214, 2016

AMERICAN SOCIETY FOR TESTING AND MATERIALS - ASTM. **Standard Practices for infrared Multivariate Quantitative Analysis**: ASTM E 1655-05-2012. United States: ASTM, 2012.

AMMARI, F.; REDJDAL, L.; RUTLEDGE, D. N. Detection of orange juice frauds using front-face fluorescence spectroscopy and independent components analysis. **Food Chemistry**, v. 168, p. 211-217, 2015.

ANDERSEN, C. M.; BRO, R. Variable selection in regression—a tutorial. **Journal of Chemometrics**, v. 24, p. 728-737, 2010.

ASADPOOR, M.; ANSARIN, M.; NEMATI, M. Amino acid profile as a feasible tool for determination of the authenticity of fruit juices. **Advanced Pharmaceutical Bulletin**, v. 4, p. 359, 2014.

ASHURST, P. R. **Chemistry and technology of soft drinks and fruit juices**. 2^a. ed. Hereford: Wiley-Blackwell, 2005. 396 p.

ASHURST, P. R. **Chemistry and technology of soft drinks and fruit juices**. John Wiley & Sons. 3^a. ed. Hereford: Wiley-Blackwell, 2016. 424 p.

ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DE REFRIGERANTES E DE BEBIDAS NÃO ALCOÓLICAS - ABIR. Dados de mercado 2010-2014. 2013.

Disponível em: <<http://abir.org.br/2015/09/01/dados-de-mercado-2010-2014/>>
Acesso em: 28 out 2015.

ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DE REFRIGERANTES E DE BEBIDAS NÃO ALCOÓLICAS - ABIR. Dados. 2014. Disponível em: <<https://abir.org.br/o-setor/dados/>> Acesso em: 31 out 2018.

ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DE REFRIGERANTES E DE BEBIDAS NÃO ALCOÓLICAS - ABIR. Bruna Torres. Secretaria Geral e Comunicação. Mensagem recebida de btorres@abir.org.br em dezembro de 2015

BARKER, M.; RAYENS, W. Partial least squares for discrimination. **Journal of Chemometrics**, v. 17, p. 166-173, 2003.

BARROS NETO, B. D.; SCARMINIO, I. S.; BRUNS, R. E. 25 anos de quimiometria no Brasil. **Química Nova**, v. 29, p. 1401-1406, 2006.

BATES, R. P.; MORRIS, J. R.; CRANDALL, P. G. **Principles and practices of small-and medium-scale fruit juice processing**, Food and Agriculture Organization of the United Nations, 2001.

BEULLENS, K. *et al.* The electronic tongue and ATR–FTIR for rapid detection of sugars and acids in tomatoes. **Sensors and Actuators B: Chemical**, v. 116, p. 107-115, 2006.

BEVILACQUA, M. *et al.* Classification and class-modelling. In: MARINI, F. **Data handling in science and technology**: v.28, 2013. p.171-233.

BOGGIA, R. *et al.* A screening method based on UV–Visible spectroscopy and multivariate analysis to assess addition of filler juices and water to pomegranate juices. **Food Chemistry**, v. 140, p. 735-741, 2013.

BONA, E. *et al.* Support vector machines in tandem with infrared spectroscopy for geographical classification of green arabica coffee. **LWT-Food Science and Technology**, v. 76, p. 330-336, 2017.

BOTELHO, B. G.; MENDES, B. A.; SENA, M. M. Development and analytical validation of robust near-infrared multivariate calibration models for the quality inspection control of mozzarella cheese. **Food Analytical Methods**, v. 6, p. 881-891, 2013.

BOTELHO, B. G. *et al.* Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. **Food Chemistry**, v. 181, p. 31-37, 2015.

BRASIL. Decreto nº 6.871, de 4 de junho de 2009. Regulamenta a Lei nº 8.918, de 14 de julho de 1994, que dispõe sobre a padronização, a classificação, o registro, a inspeção, a produção e a fiscalização de bebidas. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 4 jun. 2009.

_____. Decreto-Lei nº 2.848, de 7 de dezembro de 1940. Código penal. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, dez. 1940. Seção 1, p. 23911.

_____. Lei nº 8.078, de 11 de setembro de 1990. Dispõe sobre a proteção do consumidor e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 12 set. 1990.

_____. Lei nº 8.918, de 14 de julho de 1994. Dispõe sobre a padronização, a classificação, o registro, a inspeção, a produção e a fiscalização de bebidas, autoriza a criação da Comissão Intersectorial de Bebidas e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, julho. 1994.

BRERETON, R. G. **Applied Chemometrics for Scientists**. England: John Wiley & Sons, 2007. 379 p.

_____. Pattern recognition in chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 149, p. 90-96, 2015.

BRO, R. Multivariate calibration: What is in chemometrics for the analytical chemist? **Analytica Chimica Acta**, v. 500, p. 185-194, 2003.

BROADHURST, D. *et al.* Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry. **Analytica Chimica Acta**, v. 348, p. 71-86, 1997.

BROWN, C. D.; GREEN, R. L. Critical factors limiting the interpretation of regression vectors in multivariate calibration. **TrAC - Trends in Analytical Chemistry**, v. 28, p. 506-514, 2009.

BUREAU, S. *et al.* MApplication of ATR-FTIR for a rapid and simultaneous determination of sugars and organic acids in apricot fruit. **Food Chemistry**, v. 115, p. 1133-1140, 2009.

BUREAU, S. *et al.* Effect of sample preparation on the measurement of sugars, organic acids, and polyphenols in apple fruit by mid-infrared spectroscopy. **Journal of Agricultural and Food Chemistry**, v. 60, p. 3551-3563, 2012.

BYLESJÖ, M. *et al.* OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. **Journal of Chemometrics**, v. 20, p. 341-351, 2006.

CABANISS, S.; LEENHEER, J.; MCVEY, I. Aqueous infrared carboxylate absorbances: aliphatic di-acids. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 54, p. 449-458, 1998.

CALLAO, M. P.; RUISÁNCHEZ, I. An overview of multivariate qualitative methods for food fraud detection. **Food Control**, v. 86, p. 283-293, 2018.

CANÇADO JÚNIOR, F. L.; ESTANISLAU, M. L. L.; PAIVA, B. D. Aspectos econômicos da cultura do maracujá. **Informe Agropecuário**, v. 21, p. 10-17, 2000.

CHITARRA, M.; CHITARRA, A. **Pós-colheita de frutas e hortaliças: fisiologia e manuseio**. 2. ed. Lavras: Universidade Federal de Lavras, 2005. 783 p.

CHONG, I.G.; JUN, C.H. Performance of some variable selection methods when multicollinearity is present. **Chemometrics and Intelligent Laboratory Systems**, v. 78, p. 103-112, 2005.

COATES, J. Interpretation of infrared spectra, a practical approach. **Encyclopedia of Analytical Chemistry**. 1^a. ed. Chichester: John Wiley & Sons, 2000. 24 p.

COLNAGO, L. A.; ANDRADE, F. D. D. RMN no domínio do tempo: fundamentos e aplicações offline e online. In: RESENDE, R. R. **Biotecnologia Aplicada à Agro&indústria. Fundamentos e aplicações**. 1^a. ed. São Paulo: Blucher, 2016, v.4, p.440-468.

COSTA FILHO, P. A.; POPPI, R. J. Algoritmo genético em química. **Química Nova**, v. 22, p. 405, 1999.

COZZOLINO, D. The role of vibrational spectroscopy as a tool to assess economically motivated fraud and counterfeit issues in agricultural products and foods. **Analytical Methods**, v. 7, p. 9390-9400, 2015.

CUNY, M. *et al.* Fruit juice authentication by ¹H NMR spectroscopy in combination with different chemometrics tools. **Analytical and Bioanalytical Chemistry**, v. 390, p. 419-427, 2008.

DANDAGE, K.; BADIA-MELIS, R.; RUIZ-GARCÍA, L. Indian perspective in food traceability: A review. **Food Control**, v. 71, p. 217-227, 2017.

DAVENEL, A. *et al.* NMR relaxometry as a potential non-invasive routine sensor for characterization of phenotype in *Crassostrea gigas*. **Aquaculture**, v. 291, p. 74-77, 2009.

DEFERNEZ, M.; WILSON, R. H. Mid-infrared spectroscopy and chemometrics for determining the type of fruit used in jam. **Journal of the Science of Food and Agriculture**, v. 67, p. 461-467, 1995.

DI ANIBAL, C. V.; CALLAO, M. P.; RUISÁNCHEZ, I. ¹H NMR variable selection approaches for classification. A case study: The determination of adulterated foodstuffs. **Talanta**, v. 86, p. 316-323, 2011.

DOWNEY, G. Food and food ingredient authentication by mid-infrared spectroscopy and chemometrics. **TrAC - Trends in Analytical Chemistry**, v. 17, p. 418-424, 1998.

EDELMANN, A. *et al.* Rapid method for the discrimination of red wine cultivars based on mid-infrared spectroscopy of phenolic wine extracts. **Journal of Agricultural and Food Chemistry**, v. 49, p. 1139-1145, 2001.

EL DARRA, N. *et al.* Food fraud detection in commercial pomegranate molasses syrups by UV–VIS spectroscopy, ATR-FTIR spectroscopy and HPLC methods. **Food Control**, v. 78, p. 132-137, 2017.

ELLIS, D. I. *et al.* Rapid, high-throughput, and quantitative determination of orange juice adulteration by Fourier-transform infrared spectroscopy. **Analytical Methods**, v. 8, p. 5581-5586, 2016.

ESPERANCINI, M. S. T. Mercado brasileiro de bebidas. In: VENTURINI FILHO, W. G. **Tecnologia de bebidas: matéria-prima, BPF/APPCC, legislação e mercado**. São Paulo: Edgard Blücher, 2005. cap. 2, p.21-49.

ESSLINGER, S.; RIEDL, J.; FAUHL-HASSEK, C. Potential and limitations of non-targeted fingerprinting for authentication of food in official control. **Food Research International**, v. 60, p. 189-204, 2013.

ESTEKI, M.; SHAHSAVARI, Z.; SIMAL-GANDARA, J. Use of spectroscopic methods in combination with linear discriminant analysis for authentication of food products. **Food Control**, v. 91, p. 100-112, 2018.

EUROPEAN COMMISSION. **Regulation nº 178/2002 of the european parliament and of the council**. Official Journal of the European Communities, 2002, L031.

EUROPEAN UNION. **Report on the food crisis, fraud in the food chain and the control thereof**. Procedure 2013/2091(INI). 2013.

FABER, N.; RAJKO, R. How to avoid over-fitting in multivariate calibration—The conventional validation approach and an alternative. **Analytica Chimica Acta**, v. 595, p. 98-106, 2007.

FERNÁNDEZ-GONZÁLEZ, A. *et al.* Easy-to-use analytical approach based on ATR–FTIR and chemometrics to identify apple varieties under Protected Designation of Origin (PDO). **Computers and Electronics in Agriculture**, v. 108, p. 166-172, 2014.

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

FERREIRA, M. M. C. **Quimiometria: conceitos, métodos e aplicações**. Campinas: Unicamp, 2015.

FLORES, D. *et al.* Prediction of Orange juice sensorial attributes from intact fruits by TD-NMR. **Microchemical Journal**, v. 128, p. 113-117, 2016.

GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. **Analytica Chimica Acta**, v. 185, p. 1-17, 1986.

GEORGOULI, K.; DEL RINCON, J. M.; KOIDIS, A. Continuous statistical modelling for rapid detection of adulteration of extra virgin olive oil using mid infrared and Raman spectroscopic data. **Food Chemistry**, v. 217, p. 735-742, 2017.

GOICOECHEA, H. C.; OLIVIERI, A. C. A new family of genetic algorithms for wavelength interval selection in multivariate analytical spectroscopy. **Journal of Chemometrics**, v. 17, p. 338-345, 2003.

GONDIM, C. S. *et al.* Detection of several common adulterants in raw milk by MID-infrared spectroscopy and one-class and multi-class multivariate strategies. **Food Chemistry**, v. 230, p. 68-75, 2017.

GUINÉ, R. Projeto de uma Indústria de Processamento de Néctares de Maçã e de Pera. **Millenium-Journal of Education, Technologies, and Health**, n. 42, p. 175-189, 2016.

GUYON, F. *et al.* $^{13}\text{C}/^{12}\text{C}$ isotope ratios of organic acids, glucose and fructose determined by HPLC-co-IRMS for lemon juices authenticity. **Food Chemistry**, v. 146, p. 36-40, 2014.

HAN, J. *et al.* PCR and DHPLC methods used to detect juice ingredient from 7 fruits. **Food Control**, v. 25, p. 696-703, 2012.

HE, J.; RODRIGUEZ-SAONA, L. E.; GIUSTI, M. M. Midinfrared spectroscopy for juice authentication rapid differentiation of commercial juices. **Journal of Agricultural and Food Chemistry**, v. 55, p. 4443-4452, 2007.

HIBBERT, D. B. *et al.* IUPAC project: A glossary of concepts and terms in chemometrics. **Analytica Chimica Acta**, v. 642, p. 3-5, 2009.

HÖSKULDSSON, A. Variable and subset selection in PLS regression. **Chemometrics and Intelligent Laboratory Systems**, v. 55, n. 1-2, p. 23-38, 2001.
IRUDAYARAJ, J.; TEWARI, J. Simultaneous monitoring of organic acids and sugars in fresh and processed apple juice by Fourier transform infrared–attenuated total reflection spectroscopy. **Applied Spectroscopy**, v. 57, p. 1599-1604, 2003.

JANDRIĆ, Z. *et al.* Assessment of fruit juice authenticity using UPLC–QToF MS: a metabolomics approach. **Food Chemistry**, v. 148, p. 7-17, 2014.

KAROUI, R.; PIERNA, J. F.; DUFOUR, E. Spectroscopic technique: mid-infrared (MIR) and Fourier transform mid-infrared (FT-MIR) spectroscopies. In: SUN, D.W. **Modern Techniques for Food Authentication**, Hardbound: Academic Press, 2008, p. 27-64.

KELLY, J. D.; DOWNEY, G. Detection of sugar adulterants in apple juice using Fourier transform infrared spectroscopy and chemometrics. **Journal of Agricultural and Food Chemistry**, v. 53, p. 3281-3286, 2005.

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

KIRTIL, E. *et al.* Recent advances in time domain NMR & MRI sensors and their food applications. **Current Opinion in Food Science**, v. 17, p. 9-15, 2017.

LEAO, P. D. S. **Cultivo da videira**. 1^a. ed. Petrolina: Embrapa Semiárido, 2004.

LEARDI, R. Genetic algorithms in chemometrics and chemistry: a review. **Journal of Chemometrics**, v. 15, p. 559-569, 2001.

LEARDI, R.; NORGAARD, L. Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. **Journal of Chemometrics: A Journal of the Chemometrics Society**, v.18, p. 486-497, 2004.

LI, X. *et al.* Partial Least Squares Discriminant Analysis Model Based on Variable Selection Applied to Identify the Adulterated Olive Oil. **Food Analytical Methods**, v. 9, p. 1713-1718, 2016.

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

LÓPEZ, M. I. *et al.* Multivariate screening in food adulteration: Untargeted versus targeted modelling. **Food Chemistry**, v. 147, p. 177-181, 2014.

LUCASIU, C. B.; KATEMAN, G. Genetic algorithms for large-scale optimization in chemometrics: an application. **TrAC - Trends in Analytical Chemistry**, v. 10, p. 254-261, 1991.

MANFREDI, M. *et al.* Fast classification of hazelnut cultivars through portable infrared spectroscopy and chemometrics. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 189, p. 427-435, 2018

MANLEY, M.; DOWNEY, G.; BAETEN, V. Spectroscopic technique: near-infrared (NIR) spectroscopy. In: SUN, D.W. **Modern techniques for food authentication**, Hardbound: Academic Press, 2008, p. 65-116.

MARINI, F. Classification methods in chemometrics. **Current Analytical Chemistry**, v. 6, p. 72-79, 2010.

MÁRQUEZ, C. *et al.* FT-Raman and NIR spectroscopy data fusion strategy for multivariate qualitative analysis of food fraud. **Talanta**, v. 161, p. 80-86, 2016.

MARTINS, A. R. *et al.* Discrimination of whisky brands and counterfeit identification by UV-Vis spectroscopy and multivariate data analysis. **Food Chemistry**, v. 229, p. 142-151, 2017.

MEDEIROS, A. D. **Cultivo do pessegueiro**. 1ª. ed. Pelotas: Embrapa Clima Temperado, 2002.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO - MAPA. **Instrução Normativa nº 12**, de 4 de setembro de 2003. Aprova o Regulamento Técnico para Fixação dos Padrões de Identidade e Qualidade Gerais para Suco Tropical; os Padrões de Identidade e Qualidade dos Sucos Tropicais de Abacaxi, Acerola, Cajá, Caju, Goiaba, Graviola, Mamão, Manga, Mangaba, Maracujá e Pitanga; e os Padrões de Identidade e Qualidade dos Néctares de Abacaxi, Acerola, Cajá, Caju, Goiaba, Graviola, Mamão, Manga, Maracujá, Pêssego e Pitanga. Diário Oficial [da] República Federativa do Brasil, Brasília, DF, 9 set. 2003.

_____. **Instrução Normativa nº 24**, de 08 de setembro de 2005. Aprova o Manual Operacional de Bebidas e Vinagre, onde constam o modelo de manual da qualidade, o protocolo de validação de métodos físico-químicos, o protocolo de cálculo da incerteza associada às medições e os métodos de análise de bebidas fermentadas, destiladas, não alcoólicas e de vinagre, que passam a constituir padrões oficiais para análise físico-química de bebidas e vinagre. Diário Oficial [da] República Federativa do Brasil, Brasília, DF, 20 set. 2005. Seção 1, p. 11.

_____. **Instrução Normativa nº 42** de 11 de setembro de 2013. Altera a quantidade mínima de polpa de suco de uva e de laranja em seus respectivos néctares. Diário Oficial [da] República Federativa do Brasil, Brasília, DF, 11 set. 2013. Seção 1, p. 3.

MORAES, I. V. M. **Dossiê técnico**. Produção de Polpa de Fruta Congelada e Suco de Frutas. Rio de Janeiro: Redetec, 2006. 23 p.

MOROS, J. *et al.* Determination of the energetic value of fruit and milk-based beverages through partial-least-squares attenuated total reflectance-Fourier transform infrared spectrometry. **Analytica Chimica Acta**, v. 538, p. 181-193, 2005.

NASCIMENTO, C. F. *et al.* Recent advances on determination of milk adulterants. **Food Chemistry**, v. 221, p. 1232-1244, 2017.

NASCIMENTO NETO, F. D. **Recomendações básicas para a aplicação das boas práticas agropecuárias e de fabricação na agricultura familiar**. 1. ed. Brasília: Embrapa Informação Tecnológica, 2006. 243 p.

NEVES, M. F. *et al.* **The orange juice business: A Brazilian perspective.** The Netherlands: Wageningen Academic Publishers, 2012. 175 p.

NICOLAI, B. M. *et al.* Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. **Postharvest Biology and Technology**, v. 46, p. 99-118, 2007.

NORGAARD, L. *et al.* Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. **Applied Spectroscopy**, v. 54, p. 413-419, 2000.

NUNCIO-JÁUREGUI, N. *et al.* Pomegranate juice adulteration by addition of grape or peach juices. **Journal of the Science of Food and Agriculture**, v. 94, p. 646-655, 2014.

OLIVEIRA, F. **Sistema de produção para manejo do cajueiro comum e recuperação de pomares improdutivos.** Fortaleza: Embrapa Agroindústria Tropical-Sistema de Produção, 2007.

OLIVERI, P. Class-modelling in food analytical chemistry: Development, sampling, optimisation and validation issues—A tutorial. **Analytica Chimica Acta**, v. 982, p. 9-19, 2017.

OLIVERI, P.; DOWNEY, G. Multivariate class modeling for the verification of food-authenticity claims. **TrAC - Trends in Analytical Chemistry**, v. 35, p. 74-86, 2012.

OLIVERI, P. *et al.* Partial least squares density modeling (PLS-DM)—A new class-modeling strategy applied to the authentication of olives in brine by near-infrared spectroscopy. **Analytica Chimica Acta**, v. 851, p. 30-36, 2014.

PALTRINIERI, G.; FIGUEROLA, F.; ROJAS, L. **Technical manual on small-scale processing of fruits and vegetables.** Food and Agriculture Organization of the United Nations, 1997.

PALTRINIERI, G.; FIGUEROLA, F. **Small-scale processing of native and introduced Amazonian fruits and vegetables.** Technical manual. Food and Agriculture Organization of the United Nations, 1998.

PARDO, M. A. Evaluation of a dual-probe real time PCR system for detection of mandarin in commercial orange juice. **Food Chemistry**, v. 172, p. 377-384, 2015.

PEREIRA, F. M. V.; DE SOUZA CARVALHO, A.; CABEÇA, L. F.; & COLNAGO, L. A. Classification of intact fresh plums according to sweetness using time-domain nuclear magnetic resonance and chemometrics. **Microchemical Journal**, v. 108, p. 14-17, 2013. (a)

PEREIRA, F. M. V.; PFLANZER, S. B.; GOMIG, T.; GOMES, C. L.; DE FELÍCIO, P. E.; & COLNAGO, L. A. Fast determination of beef quality parameters with time-domain nuclear magnetic resonance spectroscopy and chemometrics. **Talanta**, v. 108, p. 88-91, 2013. (b)

PEREIRA, M. E. C. *et al.* **Procedimentos pós-colheita na produção integrada de citros**. Cruz das Almas: Embrapa Mandioca e Fruticultura Tropical, 2006.

POMERANTSEV, A. L.; RODIONOVA, O. Y. Concept and role of extreme objects in PCA/SIMCA. **Journal of Chemometrics**, v. 28, p. 429-438, 2014a.

_____. On the type II error in SIMCA method. **Journal of Chemometrics**, v. 28, p. 518-522, 2014b.

POPPI, R.; SENA, M. D. Avaliação do uso de métodos quimiométricos em análise de solo. **Química Nova**, v. 23, n. 4, 2000.

PULIDO, A. *et al.* Uncertainty of results in routine qualitative analysis. **TrAC - Trends in Analytical Chemistry**, v. 22, p. 647-654, 2003.

REID, L. M.; O'DONNELL, C. P.; DOWNEY, G. Recent technological advances for the determination of food authenticity. **Trends in Food Science & Technology**, v. 17, p. 344-353, 2006.

RINKE, P. Tradition Meets High Tech for Authenticity Testing of Fruit Juices. In: DOWNEY, G. **Advances in Food Authenticity Testing**: Elsevier, 2016. p.625-665.

RÍOS-REINA, R. *et al.* ATR-FTIR as a potential tool for controlling high quality vinegar categories. **Food Control**, v. 78, p. 230-237, 2017.

RIUS, A.; P. CALLAO, M.; X. RIUS, F. Multivariate Statistical Process Control Applied to Sulfate Determination by Sequential Injection Analysis. **Analyst**, v. 122, p. 737-741, 1997.

RODIONOVA, O. Y.; OLIVERI, P.; POMERANTSEV, A. L. Rigorous and compliant approaches to one-class classification. **Chemometrics and Intelligent Laboratory Systems**, v. 159, p. 89-96, 2016.

SANTOS, P. M.; COLNAGO, L. A. Comparison among MIR, NIR, and LF-NMR techniques for quality control of jam using chemometrics. **Food Analytical Methods**, v. 11, p. 2029-2034, 2018.

SANTOS, P. M.; PEREIRA-FILHO, E. R.; COLNAGO, L. A. Detection and quantification of milk adulteration using time domain nuclear magnetic resonance (TD-NMR). **Microchemical Journal**, v. 124, p. 15-19, 2016.

SEN, I.; TOKATLI, F. Differentiation of wines with the use of combined data of UV-visible spectra and color characteristics. **Journal of Food Composition and Analysis**, v. 45, p. 101-107, 2016.

SHAH, N. *et al.* Use of attenuated total reflectance midinfrared for rapid and real-time analysis of compositional parameters in commercial white grape juice. **Journal of Agricultural and Food Chemistry**, v. 58, p. 3279-3283, 2010.

SHAO, X.; LI, Y. Classification and prediction by LF NMR. **Food and Bioprocess Technology**, v. 5, p. 1817-1823, 2012.

SHAO, X.; LI, Y. Application of low-field NMR to analyze water characteristics and predict unfrozen water in blanched sweet corn. **Food and Bioprocess Technology**, v. 6, p. 1593-1599, 2013.

SHEN, F. *et al.* Detection of Adulteration in Freshly Squeezed Orange Juice by Electronic Nose and Infrared Spectroscopy. **Czech Journal of Food Science**, v. 34, p.224-232, 2016..

SINGHAL, R. S.; KULKARNI, P. R.; REGE, D. V. Fruit and Vegetable Products. In: _____. **Handbook of Indices of Food Quality and Authenticity**: England: Woodhead Publishing, 1997. p.77-130.

SIVAKESAVA, S.; IRUDAYARAJ, J. A rapid spectroscopic technique for determining honey adulteration with corn syrup. **Journal of Food Science**, v. 66, p. 787-791, 2001.

SNYDER, A. B. *et al.* Rapid authentication of concord juice concentration in a grape juice blend using Fourier-Transform infrared spectroscopy and chemometric analysis. **Food Chemistry**, v. 147, p. 295-301, 2014.

SORIA, A. *et al.* Chromatographic Technique: Gas Chromatography (GC). In: SUN, D.W. **Modern Techniques for Food Authentication**, Hardbound: Academic Press, 2008, p. 321-359.

SOROL, N. *et al.* Visible/near infrared-partial least-squares analysis of Brix in sugar cane juice: A test field for variable selection methods. **Chemometrics and Intelligent Laboratory Systems**, v. 102, p. 100-109, 2010.

SOUZA, Scheilla Vitorino Carvalho de. **Procedimento para validação intralaboratorial de métodos de ensaio**: delineamento e aplicabilidade em análises de alimentos. 2007. 297f. Tese (Doutorado em Ciência de Alimentos) – Faculdade de Farmácia, Universidade Federal de Minas Gerais, Belo Horizonte, 2007.

SOUZA, S. V.; JUNQUEIRA, R. G. A procedure to assess linearity by ordinary least squares method. **Analytica Chimica Acta**, v. 552, p. 25-35, 2005.

SOUZA, S. V.; PINTO, C. T.; JUNQUEIRA, R. G. In-house method validation: application in arsenic analysis. **Journal of Food Composition and Analysis**, v. 20, p. 241-247, 2007.

SPINELLI, F. *et al.* Detection of addition of apple juice in purple grape juice. **Food Control**, v. 69, p. 1-4, 2016.

SPRAUL, M. *et al.* NMR-based multi parametric quality control of fruit juices: SGF profiling. **Nutrients**, v. 1, p. 148-155, 2009.

STUART, B. H. **Infrared spectroscopy: Fundamentals and applications.** England: John Wiley & Sons, 2004. 244 p.

SZYMAŃSKA, E. *et al.* Chemometrics and qualitative analysis have a vibrant relationship. **TrAC Trends in Analytical Chemistry**, v. 69, p. 34-51, 2015.

TEÓFILO, R. F.; MARTINS, J. P. A.; FERREIRA, M. Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression. **Journal of Chemometrics**, v. 23, p. 32-48, 2009.

THOMPSON, M.; ELLISON, S. L.; WOOD, R. Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC Technical Report). **Pure and Applied Chemistry**, v. 74, p. 835-855, 2002.

TODT, H. *et al.* Water/moisture and fat analysis by time-domain NMR. **Food Chemistry**, v. 96, p. 436-440, 2006.

TRESSLER, D. K.; JOSLYN, M. A. **Fruit and vegetable juice processing technology.** 1. ed. Connecticut: Avi Publishing Company, 1961. 1028 p.

VALDERRAMA, P.; BRAGA, J. W.; POPPI, R. J. Estado da Arte de Figuras de Mérito em Calibração Multivariada. **Química Nova**, v. 32, p. 1278-1287, 2009.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Variable selection, outlier detection, and figures of merit estimation in a partial least-squares regression multivariate calibration model. A case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy. **Journal of Agricultural and Food Chemistry**, v. 55, p. 8331-8338, 2007.

VARDIN, H. *et al.* Authentication of pomegranate juice concentrate using FTIR spectroscopy and chemometrics. **Food Chemistry**, v. 108, p. 742-748, 2008.

VESSMAN, J. *et al.* Selectivity in analytical chemistry (IUPAC Recommendations 2001). **Pure and Applied Chemistry**, v. 73, p. 1381-1386, 2001.

VISÃO OESTE. 2016. Alerta: fraudes em bebidas vão além de produtos alcoólicos. Disponível em: <<https://www.visaoeste.com.br/alerta-fraudes-em-bebidas-vaio-alem-de-produtos-alcoolicos/>>. Acesso em: 24 nov. 2016.

WILLIAMS, P. Implementation of near-infrared technology. In: WILLIAMS, P. & NORRIS, K.. **Near-infrared technology in the agricultural and food industries**. 2. ed. St. Paul: American Association of Cereal Chemists Inc. 2001. p. 145-169.

WILSON, R. Fourier transform mid-infrared spectroscopy for food analysis. **TrAC - Trends in Analytical Chemistry**, v. 9, p. 127-131, 1990.

WISE, B. M. *et al.* PLS_Toolbox Version 4.0 for use with MATLAB™. **Eigenvector Research, Inc**, v. 3905, 2006.

WOLD, S. Chemometrics; what do we mean with it, and what do we want from it? **Chemometrics and Intelligent Laboratory Systems**, v. 30, p. 109-115, 1995.

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

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

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

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

ZHANG, J. *et al.* Review of the current application of fingerprinting allowing detection of food adulteration and fraud in China. **Food Control**, v. 22, p. 1126-1135, 2011.

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