

UNIVERSIDADE FEDERAL DE MINAS GERAIS – UFMG
INSTITUTO DE CIÊNCIAS EXATAS – ICEX
DEPARTAMENTO DE QUÍMICA

Bruno Gonçalves Botelho

Desenvolvimento e validação de métodos para
análise direta de alimentos usando ferramentas
quimiométricas, espectroscopia no infravermelho e
imagens digitais

Belo Horizonte
2014

UFMG-ICEx/DQ.1044^a

T.473^a.

Bruno Gonçalves Botelho

Desenvolvimento e validação de métodos para análise direta de alimentos usando ferramentas quimiométricas, espectroscopia no infravermelho e imagens digitais

Development and analytical validation of direct food analysis methods using chemometrics tools, infrared spectroscopy and digital images

Tese apresentada ao Departamento de Química do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Doutor em Ciências - Química

Belo Horizonte

2014

B748d Botelho, Bruno Gonçalves
2014 Desenvolvimento e validação de métodos para análise direta de alimentos usando ferramentas quimiométricas, espectroscopia no infravermelho e imagens digitais [manuscrito] / Bruno Gonçalves Botelho. 2014.

136 f. : il.

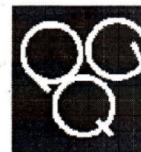
Orientador: Marcelo Martins de Sena.

Tese (doutorado) – Universidade Federal de Minas Gerais – Departamento de Química.

Inclui bibliografia.

UFMG

PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA
DEPARTAMENTO DE QUÍMICA - ICEX
31270-901 - BELO HORIZONTE - MG
TEL.: 031 - 3409-5732
FAX: 031 - 3409-5711
E-MAIL: pgquimic@qui.ufmg.br



**"Desenvolvimento e Validação de Métodos para Análise Direta de Alimentos
usando Ferramentas Quimiométricas, Espectroscopia no Infravermelho e
Imagens Digitais"**

Bruno Gonçalves Botelho

Tese aprovada pela banca examinadora constituída pelos Professores:

Prof. Marcelo Martins de Sena - Orientador
UFMG

Prof. Ronei Jesus Poppi
UNICAMP

Prof. Anselmo Elcana de Oliveira
UFG

Prof. Scheilla Vitorino Carvalho Souza Ferreira
UFMG

Prof. Leticia Malta Costa
UFMG

Belo Horizonte, 1º de outubro de 2014.

AGRADECIMENTOS

Gostaria de agradecer primeiramente à minha família, por todo apoio dispensado não só nesses quatro anos, mas ao longo de toda a minha vida acadêmica.

À Keila, minha noiva e companheira, que dividiu comigo as alegrias e os perengues enfrentados nesse doutorado. E ao Bolinha, que, nos privou de algumas horas de sono, mas a cada sorriso banguela, faz tudo valer a pena.

Ao Marcelo, que além de orientar e transmitir conhecimento, sempre se mostrou preocupado com minha formação moral e profissional com relação ao meio acadêmico.

Aos companheiros de grupo, incluindo a Cris, que dividiram comigo todos os momentos alegres e alguns momentos tensos, algumas roubadas em congressos, e que também contribuíram para minha formação profissional e crescimento como pessoa.

Ao IMA, por permitir a utilização do seu espaço, equipamentos e infra-estrutura, e em especial à Bruna, que tanto nos auxiliou no início desde o início desse projeto

Aos professores DQ, em especial aos professores da Analítica, por todo o conhecimento transmitido.

Aos funcionários do DQ, em especial às funcionárias da secretaria de pós graduação, que sempre estavam dispostas a nos ajudar.

Aos amigos de longa data, que, mesmo não entendendo nada de quimiometria, sempre estavam dispostos a tomar cerveja!

E aos órgãos de fomento, que tornaram a realização desta tese possível.

Não! Tentar não!
Faça.
Ou não faça.
Tentativa não há.
Mestre Yoda

Sumário

1. Introdução	1
2. Revisão bibliográfica	3
2.1 INTERAÇÃO LUZ/MATÉRIA	3
2.2 ESPECTROSCOPIA VIBRACIONAL	3
2.2.1 OSCILAÇÃO HARMÔNICA E ANARMÔNICA	4
2.3 ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO (NIRS)	6
2.4 ESPECTROSCOPIA NO INFRAVERMELHO MÉDIO (MIRS)	7
2.5 ANÁLISE DE IMAGENS DIGITAIS	10
2.6 QUIMIOMETRIA	12
2.6.1 CALIBRAÇÃO MULTIVARIADA	12
2.6.2 MÍNIMOS QUADRADOS PARCIAIS (PLS)	13
2.6.3 N-PLS.....	14
2.6.5 PLS-DA	17
2.6.6 QUIMIOMETRIA APLICADA À ANÁLISE DE ALIMENTOS.....	18
2.7 VALIDAÇÃO ANALÍTICA MULTIVARIADA	19
2.7.1 DETECÇÃO DE AMOSTRAS ANÔMALAS (OUTLIERS).....	19
2.7.2 SINAL ANALÍTICO LÍQUIDO (NAS) E CURVAS PSEUDOUNIVARIADAS.....	20
2.7.3 FIGURAS DE MÉRITO (FOM)	21
2.7.4 VALIDAÇÃO QUALITATIVA MULTIVARIADA.....	29
2.7.5 MONITORAMENTO PÓS-MODELO	32
3. Modelos para o controle fiscal de queijos utilizando espectroscopia NIR e PLS	33
3.1 DETERMINAÇÃO DE UMIDADE E GORDURA EM QUEIJOS MUÇARELA	34
3.1.2 OBJETIVOS.....	36
3.1.3 MATERIAIS E MÉTODOS.....	37
3.1.4 RESULTADOS E DISCUSSÃO	38
3.2. DETERMINAÇÃO DE UMIDADE EM QUEIJOS MINAS ARTESANAL UTILIZANDO NIR E PLS	53
3.2.1 OBJETIVOS.....	54
3.2.2 MATERIAL E MÉTODOS.....	55
3.2.3 RESULTADOS E DISCUSSÃO	55
3.3 CONCLUSÃO	64
4. Desenvolvimento de um método de triagem para a detecção simultânea de cinco adulterantes em leite cru utilizando infravermelho médio e PLS-DA	66
4.1 INTRODUÇÃO	66

4.2 OBJETIVOS	68
4.3 MATERIAIS E MÉTODOS	68
4.3.1 EQUIPAMENTOS E SOFTWARES	68
4.3.2 AMOSTRAS E REAGENTES.....	68
4.3.3 PROCEDIMENTO.....	68
4.4 RESULTADOS E DISCUSSÃO	71
4.4.1 ESPECTROS DE INFRAVERMELHO	71
5.4.2 MODELOS PLS-DA	73
4.4.3 VALIDAÇÃO QUALITATIVA MULTIVARIADA.....	78
4.5 CONCLUSÃO	80
5. Determinação do corante amarelo crepúsculo em bebidas não alcoólicas utilizando análise multivariada de imagens	80
5.1 INTRODUÇÃO	80
5.2 OBJETIVOS	82
5.3 MATERIAIS E MÉTODOS	83
5.3.1 EQUIPAMENTOS E SOFTWARES	83
5.3.2 AMOSTRAS	83
5.3.3 PROCEDIMENTO.....	83
5.3.4 ANÁLISES CROMATOGRÁFICAS	85
5.4 RESULTADOS E DISCUSSÃO	86
5.4.1 MODELO PLS.....	86
5.4.2 VALIDAÇÃO ANALÍTICA MULTIVARIADA	88
5.4.3 B-TI	92
5.5 CONCLUSÕES	93
6. Telefone celular como instrumento de aquisição de imagens para modelos multivariados: Determinação do corante vermelho 40 em balas duras	94
6.2 OBJETIVOS	97
6.3 MATERIAIS E MÉTODOS	97
6.3.1 EQUIPAMENTOS E SOFTWARES	97
6.3.2 AMOSTRAS	97
6.3.3 PROCEDIMENTO.....	98
6.4 RESULTADOS E DISCUSSÃO	100
6.4.1 MODELO PLS/RGB	100
6.4.2 MODELO N-PLS/FFT	101
6.4.3 VALIDAÇÃO ANALÍTICA MULTIVARIADA	101

6.4.4 B-TI	104
6.5 CONCLUSÕES	105
7 Conclusões Gerais	106
8 Referências Bibliográficas	107

Índice de Figuras

<i>Figura 1 – Representação esquemática dos modelos do oscilador harmônico e anarmônico¹</i>	<i>5</i>
<i>Figura 2 – Representação gráfica do sistema RGB</i>	<i>10</i>
<i>Figura 3 – Formação de tons no sistema RGB</i>	<i>11</i>
<i>Figura 4 - Decomposição de um arranjo multidimensional através do modelo N-PLS.....</i>	<i>15</i>
<i>Figura 5 - Representação esquemática da decomposição da matriz Y através do modelo NPLS</i>	<i>16</i>
<i>Figura 6 – Representação vetorial do NAS</i>	<i>20</i>
<i>Figura 7 – Butirômetro de Gerber.....</i>	<i>35</i>
<i>Figura 8 – Espectros das 123 amostras de queijo muçarela.....</i>	<i>39</i>
<i>Figura 9 - Espectros de queijo muçarela pré-processados com MSC.....</i>	<i>40</i>
<i>Figura 10 – Coeficientes de regressão para o modelo de determinação de gordura em queijo muçarela.....</i>	<i>41</i>
<i>Figura 11 - Coeficientes de regressão para o modelo de determinação de umidade em queijo muçarela</i>	<i>41</i>
<i>Figura 12 – Gráfico de resíduos vs. valores preditos para o modelo de determinação de gordura – amostras de calibração (●) e de validação (Δ)</i>	<i>44</i>
<i>Figura 13 - Gráfico de resíduos vs. valores preditos para o modelo de determinação de umidade – amostras de calibração (●) e de validação (Δ)....</i>	<i>44</i>
<i>Figura 14 – Curva de valores de referência vs. norma do NAS para o modelo de determinação de gordura - amostras de calibração (●) e de validação (Δ) .</i>	<i>46</i>
<i>Figura 15 - Curva de valores de referência vs. norma do NAS para o modelo de determinação de umidade - amostras de calibração (●) e de validação (Δ)....</i>	<i>46</i>
<i>Figura 16 – Erros relativos das amostras de validação para o modelo de determinação de gordura</i>	<i>47</i>
<i>Figura 17 - Erros relativos das amostras de validação para o modelo de determinação de umidade</i>	<i>48</i>
<i>Figura 18 – Efeitos do teste de robustez, de acordo com um planejamento 2²</i>	<i>50</i>

<i>Figura 19 - Erro amostral específico de predição das amostras de validação para o modelo de determinação de gordura - valores de referência (●) e valores preditos (*).....</i>	<i>51</i>
<i>Figura 20 - Erro amostral específico de predição das amostras de validação para o modelo de determinação de umidade - valores de referência (●) e valores preditos (*).....</i>	<i>51</i>
<i>Figura 21 – Carta de controle para monitorar a estabilidade do modelo de determinação de gordura.....</i>	<i>52</i>
<i>Figura 22 - Carta de controle para monitorar a estabilidade do modelo de determinação de umidade.....</i>	<i>53</i>
<i>Figura 23 – Espectros das amostras de queijo minas artesanal.....</i>	<i>56</i>
<i>Figura 24 – Espectros de queijo minas artesanal pré-processados com MSC.....</i>	<i>56</i>
<i>Figura 25 – Coeficientes de regressão para o modelo de determinação de umidade em queijo minas artesanal.....</i>	<i>57</i>
<i>Figura 26 - Gráfico de resíduos vs. valores preditos para o modelo de determinação de umidade – amostras de calibração (●) e de validação (Δ)....</i>	<i>60</i>
<i>Figura 28 - Erros relativos das amostras de validação para o modelo de determinação de umidade.....</i>	<i>61</i>
<i>Figura 29 – Erro amostral específico de predição para as amostras de validação – valores de referência (●) e valores preditos (*).....</i>	<i>62</i>
<i>Figura 30 – Carta de controle da estabilidade do modelo para determinação de umidade em queijo minas artesanal.....</i>	<i>63</i>
<i>Figura 31 – Espectros de infravermelhos de todas as amostras de leite cru bovino (amostras brancas, adulteradas com um único adulterantes e adulteradas com as misturas de adulterantes).....</i>	<i>71</i>
<i>Figura 32 – Espectro de infravermelho médio de (a) amostra branca; (b) adulteração com água (10,0 % p/v); (c) adulteração com amido (10,0 % p/v); (d) adulteração com citrato de sódio (10,0 % p/v); (e) adulteração com formol (10,0 % p/v); (f) adulteração com sacarose (10,0 % p/v). A região espectral utilizada para a construção dos modelos está delimitada pelas linhas tracejadas.....</i>	<i>73</i>
<i>Figura 33 – Gráficos de previsão para adulterações com (a) água; (b) amido; (c) citrato de sódio; (d) formol; (e) sacarose. ▽ - Amostras Brancas; ◁ - adulteração com água; ◇ - Adulteração com amido; ○ - Adulteração com citrato; △ - Adulteração com formol; □ - Adulteração com sacarose. Marcadores cheios – Amostras de treinamento; Marcadores vazios – Amostras teste.....</i>	<i>76</i>
<i>Figura 34 – VIP escores para os modelos de detecção de adulteração de leite cru com (a) água; (b) amido; (c) citrato; (d) formol; (e) sacarose.....</i>	<i>77</i>
<i>Figura 35– Estrutura química do Amarelo Crepúsculo.....</i>	<i>82</i>
<i>Figura 36 – Escâner CanoScan LiDE 110 com o anteparo.....</i>	<i>84</i>
<i>Figura 37 – Imagem adquirida de uma amostra de refrigerante.....</i>	<i>85</i>

<i>Figura 38– Representação esquemática da decomposição matemática das imagens.....</i>	<i>85</i>
<i>Figura 40– Coeficientes de regressão do modelo PLS para determinação de AC em bebidas não alcoólicas</i>	<i>87</i>
<i>Figura 41– Gráfico de valores de referência vs. valores preditos. Calibração (quadrados) e validação (triângulo).....</i>	<i>90</i>
<i>Figura 42– Intervalos de confiança para a predição das amostras de validação, estimados com base no EAEP. Valores preditos (triângulos) e de referência (círculos).....</i>	<i>92</i>
<i>Figura 43 - β-TI estimados para o modelo (linhas azuis) e erros máximos relativos definidos pelo guia de validação brasileiro (linhas vermelhas)</i>	<i>93</i>
<i>Figura 44 – Estrutura molecular do vermelho 40.....</i>	<i>96</i>
<i>Figura 45 – Celular Motorola RAZR, empregado como instrumento para a aquisição das imagens.....</i>	<i>97</i>
<i>Figura 46 – Câmara de aquisição de imagens (vista superior fechada).....</i>	<i>98</i>
<i>Figura 49 – Imagem obtidas pelo celular de uma das amostras de bala analisadas</i>	<i>100</i>
<i>Figura 50 – Gráfico de valores preditos vs. valores de referência para o modelo de determinação do corante vermelho 40 em balas duras.....</i>	<i>102</i>
<i>Figura 51 – Erros relativos das amostras do conjunto de validação (Limites estabelecidos pela legislação em vermelho).....</i>	<i>104</i>

Índice de tabelas

<i>Tabela 1 - Otimização dos modelos PLS pela detecção de amostras anômalas (modelos finais estão destacados em negrito)</i>	<i>42</i>
<i>Tabela 2 - Figuras de mérito, estimadas para os modelos PLS de determinação de gordura e umidade</i>	<i>45</i>
<i>Tabela 3 - Otimização dos modelos PLS para determinação de umidade em queijo minas artesanal pela detecção de amostras anômalas (modelo final está destacado em negrito).....</i>	<i>57</i>
<i>Tabela 4 - Figuras de mérito estimadas para o modelo PLS de determinação de umidade em queijo minas artesanal.....</i>	<i>59</i>
<i>Tabela 5– Planejamento de mistura para a adulteração de leite.....</i>	<i>69</i>
<i>Tabela 6– Número de VLs, variâncias explicadas e erros de classificação de validação cruzada (ECVC) para os modelos PLS-DA desenvolvidos</i>	<i>74</i>

<i>Tabela 7– FOM estimadas para os modelos de detecção de adulterantes em leite cru bovino</i>	78
<i>Tabela 8 - Resultados da otimização do modelo PLS através da detecção de outliers (modelo final em negrito)</i>	88
<i>Tabela 9– Parâmetros para avaliação das FOM no modelo proposto para a determinação de AC em bebidas não alcoólicas</i>	89
<i>Tabela 10 - Resultados da otimização do modelo PLS através da detecção de outliers (modelo final em negrito)</i>	101
<i>Tabela 11 - Parâmetros para avaliação das FOM no modelo proposto para a determinação de Vermelho 40 em balas duras</i>	103

Lista de siglas e abreviações

- AC – Amarelo Crepúsculo
- ANN – Redes Neurais Artificiais (*Artificial Neural Networks*)
- ATR – Reflectância Total Atenuada (*Attenuated Total Reflectance*)
- BF – Teste de Brown-Forsythe
- CLS – Mínimos Quadrados Clássicos (*Classical Least Squares*)
- DPR – Desvio Padrão Relativo
- DRIFTS – Espectroscopia com Reflectância Difusa (*Diffuse Reflectance Spectroscopy*)
- DW – Teste de Durbin-Watson
- EAEP – Erro Amostral Específico de Predição
- ECVC – Erro de Classificação da Validação Cruzada
- FFT – Transformada Rápida de Fourier (*Fast Fourier Transform*)
- FOM – Figuras de mérito (*Figures of Merit*)
- ICS – *International Chemometrics Society*
- IMA – Instituto Mineiro de Agropecuária
- LAFQ – Laboratório de Análises Físico-químicas
- MAPA – Ministério da Agricultura, Pecuária e Abastecimento
- MIRS – Espectroscopia no Infravermelho Média (*Mid Infrared Spectroscopy*)
- MIR – Infravermelho médio (*Mid Infrared*)
- MSC – Correção de espalhamento multiplicativo (*Multiplicative Scatter Correction*)
- NAS – Sinal analítico líquido (*Net Analyte Signal*)
- NIR – Infravermelho próximo (*Near Infrared*)
- NIRS – Espectroscopia no Infravermelho próximo (*Near Infrared Spectroscopy*)
- OSC – Correção Ortogonal de Sinal (*Orthogonal Signal Correction*)

PAS – Espectroscopia Fotoacústica (*Photoacoustic Spectroscopy*)
PCR – Regressão por componentes principais (*Principal Components Regression*)
PLS – Mínimos quadrados parciais (*Partial Least Squares*)
RJ – Teste de Ryan-Joiner
RLM – Regressão Linear Múltipla
RMSEC – Raiz do erro quadrático médio de calibração (*Root Mean Square Error of Calibration*)
RMSECV – Raiz do erro quadrático médio de validação cruzada (*Root Mean Square Error of Cross Validation*)
RMSEP – Raiz do erro quadrático médio de predição (*Root Mean Square Error of prediction*)
RPD – Relação de desempenho do desvio (*Residual Prediction Deviation*)
SNV – Padronização pelo desvio padrão (Standard Normal Variate)
SDV – Desvio padrão dos erros de validação (*Standart Deviation of Validation Errors*)
SEL – Seletividade
SEN – Sensibilidade
VIM –Vocabulário Internacional de Metrologia
VIP Scores – Escores de importância de projeção de variáveis (*Variable Importance Projection Scores*)
VL – Variáveis latentes

RESUMO

Esta tese teve como principal objetivo o desenvolvimento de metodologias analíticas empregando ferramentas quimiométricas, espectroscopia no infravermelho (médio – MIRS - e próximo – NIRS) e análise de imagens digitais obtidas com equipamentos de baixo custo para o controle de qualidade de produtos alimentícios. Em todas as aplicações desenvolvidas foi dada ênfase à validação analítica multivariada, um aspecto de crescente importância visando a aplicação prática de métodos quimiométricos. Todos os métodos desenvolvidos foram validados buscando harmonização com guias e normas nacionais e internacionais. No capítulo 3 é mostrado o desenvolvimento de métodos para a determinação de parâmetros de qualidade de queijos, em condições reais de controle fiscal, num laboratório estatal. Modelos PLS foram construídos utilizando espectros NIR para determinação dos teores de umidade e gordura em queijos muçarela e umidade em queijos minas. A estabilidade dos modelos gerados foi avaliada ao longo de um ano. O modelo para determinação de gordura em queijo muçarela apresentou erros relativos de predição variando entre -8,7 e 8,1%. Já o modelo de umidade apresentou erros menores, entre -3,5 e 2,9%. O modelo de determinação de umidade em queijo minas apresentou erros entre -6,5 e 9,8%. No capítulo 4 é mostrado o desenvolvimento de um método de triagem qualitativo para a detecção simultânea de cinco adulterantes em leite cru utilizando MIRS. Água, amido, sacarose, formol e citrato de sódio foram escolhidos como adulterantes em função da sua presença em amostras de leite adulteradas encontradas nas últimas operações da Polícia Federal. Modelos PLS-DA foram construídos para a identificação dos cinco adulterantes individualmente, mesmo em amostras contendo misturas de dois ou mais deles. Todos os modelos, com exceção do modelo para determinação de amido apresentaram taxa de falso negativo menor que 5%. No capítulo 5 é relatado o desenvolvimento de um método para determinação do corante artificial amarelo crepúsculo em refrigerantes e bebidas isotônicas utilizando imagens digitais obtidas em um escâner de mesa comercial. O modelo apresentou erros de predição entre -6,2 e 9,0%. Este método também foi validado, mostrando que histogramas de frequência RGB obtidos de imagens digitais são capazes de fornecer respostas adequadas à sua utilização como sinal analítico, com uma

série de vantagens, como rapidez e simplicidade de análise. No capítulo 6, uma aplicação similar à do capítulo 5, porém empregando um telefone celular como equipamento de aquisição de imagens, foi desenvolvida para a quantificação do corante vermelho 40 em balas duras. Este modelo apresentou erros elevados, entre -24,8 e 39,5%, porém a grande maioria (mais de 80%) das amostras possui erros entre -20 e +10%.

Palavras-chave: Quimiometria, validação, análise de alimentos, espectroscopia no infravermelho, análise de imagens

ABSTRACT

The main objective of this thesis was the development of methods using chemometric tools, infrared spectroscopy (near and mid) and digital image analysis obtained using low cost equipment for foodstuff quality control. In all applications, multivariate analytical validation was emphasized, a subject with growing importance in chemometrics. All the proposed methods were validated searching for an harmonization um national and international validation guides. Chapter 3 presents the development of methods for the determination of quality parameters in cheese, in a real legal laboratory condition, in an official laboratory. PLS models were built using NIR spectra for the determination of moisture and fat in mozzarella cheese and moisture in minas cheese. The stability of the proposed models was evaluated over a year. The fat determination model for mozzarella presented erros between -8.7% and 8.1%, and the moisture determination model for the same cheese type presented smaller errors (-3.5% – 2.9%). The moisture determination model for minas chesse presented erros ranging from -6.5% to 9.8%. Chapter 4 present the development of a screening method for the simultaneous detection of five adulterants in raw milk using MIRS. Water, starch, sucrose, formaldehyde and sodium citrate were chosen as adulterants due to its presence in adulterated milk samples recently discovered in the last Federal Police operations. PLS-DA models were built for the individual identification of the five adulterants, even in samples that contained two or more adulterants. All the the models presented a false negative rate below 5%, been the starch model the only exception. Chapter 5 describes the deveploment of a method for the determination of the artificial dye sunset yellow in beverages using digitais images obtained from a commercial flatbed scanner. The model presented prediction errors varying from -6.2 to 9.0%. This method was also validated, proving that RGB histograms obtained from digital images are able to be used as analytical data, with a series of advantages, such as quickness and simplicity. In chapter 6, a similar application to the one presented in chapter 5 was developed, but using a cellphone to obtain images of hard candies, and this images were used to quantify the concentration of the allura red dye. This model presented higher errors than the ones seen in the other models in this thesis

(from -24.8% to 39.5%), but the great majority of the errors (80%) were between -20% and 10%.

Keywords: Chemometrics, validation, food analysis, infrared spectroscopy, image analysis

1. Introdução

A quantificação de compostos e a determinação de parâmetros de qualidade de produtos alimentícios sempre envolveram processos analíticos laboriosos, principalmente em função da complexa composição desse tipo de matriz, o que demanda, na grande maioria das vezes, etapas de extração e pré-tratamento das amostras. Essas etapas consomem muito tempo e utilizam solventes e reagentes que, muitas vezes, são tóxicos e oferecem risco ao analista. Tudo isso corrobora para a necessidade de desenvolver métodos analíticos visando o controle de qualidade de alimentos e bebidas, que sejam mais simples, rápidos e baratos.

Desde seu início, a quimiometria encontrou na análise de alimentos um campo de aplicação com amplas possibilidades, uma vez que suas ferramentas, que permitem obter informações químicas de uma amostra com mínimo pré-tratamento e de maneira rápida, conseguem superar o maior gargalo enfrentado pelos analistas de alimentos. As técnicas espectroscópicas, em especial, adquiriram uma importância notória no desenvolvimento de métodos que aplicam quimiometria na análise de alimentos, principalmente pela facilidade de manuseio dos equipamentos, versatilidade e rapidez.

A utilização de métodos quimiométricos como metodologias oficiais de análise praticamente inexistem, devido principalmente à ausência de guias de validação que contemplem as particularidades de métodos multivariados, como por exemplo, a não existência de curvas de calibração e a capacidade de se trabalhar em condições de baixa seletividade.

Tendo em vista o exposto acima, esta tese teve como principal objetivo desenvolver e validar métodos quimiométricos para a análise de alimentos, empregando três diferentes técnicas analíticas: espectroscopias no infravermelho próximo e médio, e análise de imagens digitais geradas por equipamentos de baixo custo. Após esta introdução, o Capítulo 2 apresenta os fundamentos das técnicas analíticas e dos métodos quimiométricos utilizados, além de discutir importantes aspectos da validação analítica multivariada. O Capítulo 3 descreve o desenvolvimento e a validação analítica de métodos para a determinação de parâmetros de qualidade em queijos (umidade e gordura em

muçarela e umidade em queijo minas) utilizando espectroscopia no infravermelho próximo e calibração multivariada. Este capítulo teve a particularidade de desenvolver os modelos em condições reais de uso, através de uma parceria com um laboratório de controle fiscal do Governo de Minas Gerais. Por isto, ele também descreve um estudo de estabilidade dos modelos gerados, que empregou cartas de controle para avaliar por quanto tempo eles continuaram a fornecer resultados aceitáveis. O Capítulo 4 descreve o desenvolvimento de um método de triagem qualitativo para a detecção simultânea de cinco adulterantes típicos (água, amido, citrato de sódio, formol e sacarose) em leite bovino cru utilizando espectroscopia no infravermelho médio e classificação supervisionada. Este método também foi validado através da estimativa de figuras de méritos específicas harmonizadas para métodos qualitativos multivariados. Os Capítulos 5 e 6 descrevem o desenvolvimento de métodos para a determinação de corantes artificiais em alimentos através da utilização de imagens digitais como sinais analíticos. Para a determinação do corante amarelo crepúsculo em refrigerantes e bebidas isotônicas foi utilizado um escâner de mesa comercial e uma metodologia de calibração multivariada a partir dos histogramas RGB extraídos das imagens. Na determinação do corante vermelho 40 em balas duras foi utilizado um equipamento artesanal, montado com um telefone celular, para aquisição das imagens digitais. Nesta aplicação, a metodologia utilizada no capítulo anterior não forneceu resultados satisfatórios, devido à presença de variações texturais nas superfícies das balas. Foi então necessário o uso de uma metodologia mais complexa, envolvendo calibração multidimensional, após a conversão das imagens utilizando transformada de Fourier.

2. Revisão bibliográfica

2.1 Interação luz/matéria

A interação entre luz e matéria tem despertado a curiosidade dos homens ao longo da história. Por volta de 130 a.C, Ptolomeu descreveu a refração da luz por uma gama de materiais transparentes e, em 1305, Von Freiburg simulou a estrutura do arco-íris utilizando esferas de vidro preenchidas com água. Por todo o século XVIII, a interação luz-matéria foi estudada, e pesquisadores como Snell, Huyguens, Newton, Bradley e Priestly formularam leis sobre a refração e a reflexão da luz, assim como a proposição de sua natureza dual (partícula/onda). Porém, a descoberta da radiação no infravermelho é creditada ao cientista inglês William Herschell,¹ em 1800. A teoria moderna sobre a luz, baseada na ondulação de um campo eletromagnético, foi desenvolvida por JC Maxwell (1831 – 1879).²

Os primeiros experimentos modernos utilizando técnicas espectroscópicas foram realizados por W. W. Coblentz, em 1905. Coblentz obteve o espectro de 19 compostos entre 800 e 2800 nm, com um espectrômetro equipado com cristal de quartzo e um detector caseiro. A partir destes primeiros espectros, foi possível concluir que cada substância possui um espectro característico e que grupos químicos idênticos presentes em moléculas diferentes exibem absorções nos mesmos comprimentos de onda.²

O desenvolvimento de métodos analíticos utilizando a NIRS (*Near Infrared Spectroscopy*, Espectroscopia no Infravermelho próximo) só começou em 1949, quando o Departamento da Agricultura dos Estados Unidos (USDA) lançou um projeto para avaliar a qualidade de ovos. O primeiro estudo relacionado a análise quantitativa de um composto foi publicada em 1962, por Hart *et al.*³, descrevendo um método para a determinação de umidade em sementes utilizando espectroscopia no infravermelho próximo.²

2.2 Espectroscopia vibracional

A absorção na região do infravermelho induz os movimentos vibracionais das moléculas. Essas absorções têm sido utilizadas de forma qualitativa e

quantitativa para diversas análises e para a atribuição das bandas de absorção a grupos químicos específicos. De forma geral, o movimento vibracional de dois átomos de uma molécula diatômica pode ser considerado como a compressão e extensão de uma mola, na qual os átomos se atraem e se repelem.¹

2.2.1 Oscilação harmônica e anarmônica

A oscilação harmônica é o modelo mais simples de vibração, no qual massas vibrantes m_1 e m_2 variam a distância internuclear em até 10%. Nesse caso, a lei de Hooke é aplicável, e a energia potencial, V , pode ser representada como

$$V = \frac{1}{2}k(r - r_e)^2 = \frac{1}{2}kq^2 \quad (1)$$

onde k é a força da ligação, r é a distância internuclear durante a vibração, r_e é a distância internuclear de equilíbrio. A curva da energia potencial de tal oscilador é parabólica e simétrica de acordo com r_e . Este modelo leva a uma frequência vibracional ν_0

$$\nu_0 = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad (2)$$

onde a massa reduzida m é dada por

$$m = \frac{m_1 m_2}{m_1 + m_2} \quad (3)$$

A partir das equações 2 e 3 é possível verificar que a frequência vibracional depende da estrutura do composto analisado, e essa é a base da grande utilização da espectroscopia no infravermelho para a elucidação de compostos.¹

A interação da radiação com a molécula, entretanto, só é possível se o vetor elétrico da radiação oscilar com a mesma frequência que o momento de dipolo molecular, μ . Então, uma vibração só é ativada no infravermelho se o momento de dipolo molecular for modulado pela vibração e se $\partial\mu/\partial q$ for diferente de zero. O requerimento de uma mudança no momento de dipolo durante a vibração faz do infravermelho médio (MIR) especialmente sensível a grupos funcionais polares.¹

Entretanto, o modelo do oscilador harmônico não é adequado a maiores amplitudes de vibrações, devido às forças repulsivas entre os átomos vibrantes e a possibilidade de dissociação, quando a ligação é muito estendida. Diferente da oscilação harmônica, os níveis de energia não são equidistantes no modelo anarmônico, e a regra de seleção estrita ($\Delta n = \pm 1$) é expandida para mais de um nível energético. Com isso, a curva de energia potencial é representada por uma função de Morse assimétrica (Figura 1).¹

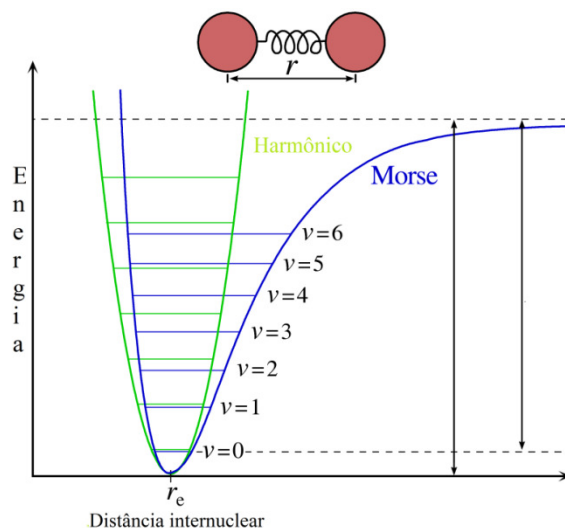


Figura 1 – Representação esquemática dos modelos do oscilador harmônico e anarmônico¹

Normalmente, uma molécula não linear contendo N átomos terá $3N-6$ graus de liberdade vibracionais, enquanto uma molécula linear possui $3N-5$. O número de graus de liberdade vibracional representa o número de frequências vibracionais da molécula. Para uma determinada molécula, um modo de vibração normal corresponde ao movimento atômico no qual todos os átomos se movem com a mesma frequência, mas com amplitudes diferentes. Adicionalmente às vibrações normais, transições correspondentes a $\Delta n = \pm 2, \pm 3 \dots$ são agora permitidas e são chamadas de sobretons, que ocorrem na região do infravermelho próximo.¹

Além dos sobretons, combinações de diferentes transições vibracionais (soma e diferença de vibrações fundamentais) também podem ser observadas. Entretanto, a probabilidade dessas transições diminui significativamente com a sua ordem, e geralmente a banda de absorção correspondente a transições de sobreton ou combinação tem intensidade inferior a suas análogas fundamentais.

Ao contrário do MIR, a região do NIR contém quase que exclusivamente bandas de absorção que podem ser atribuídas às vibrações de sobretons e combinações. Infelizmente, a sobreposição dessas bandas de sobretons e combinações diminuem significativamente a especificidade do NIR, o que é uma das principais razões pelas quais esta região foi subestimada por espectroscopistas durante muito tempo.¹

2.3 Espectroscopia no Infravermelho próximo (NIRS)

A espectroscopia no infravermelho próximo (NIRS, do inglês *near infrared spectroscopy*) é um tipo de espectroscopia vibracional, a qual emprega energia do fóton ($h\nu$) no intervalo de $2,65 \times 10^{-19}$ a $7,96 \times 10^{-20}$ J, que corresponde à faixa de comprimento de onda de 750 a 2500 nm (números de onda: 13300 a 4000 cm^{-1}). Esta faixa de energia é maior do que a necessária para promover as moléculas somente para seu estado vibracional excitado de menor energia (através de uma transição vibracional fundamental) e menor que os valores necessários para a excitação de elétrons em moléculas (com exceção de alguns compostos de terras raras).⁴

A combinação da energia da radiação com a diferença de energia entre dois níveis vibracionais provoca uma resposta seletiva do sistema molecular à radiação incidente. Isso significa que em determinado comprimento de onda, uma frequência correspondente é absorvida. Esta figura complexa da intensidade de absorção *versus* comprimento de onda constitui o espectro de absorção de uma substância ou amostra. O principal objetivo do uso da espectroscopia NIR é sondar uma amostra, a fim de adquirir dados qualitativos e/ou quantitativos provenientes da interação de ondas eletromagnéticas no infravermelho próximo com seus componentes.⁴

Os métodos analíticos resultantes do uso da região espectroscópica NIR propiciam análises rápidas (um minuto ou menos, por exemplo), não destrutivas, não invasivas, com alta penetração do feixe de radiação de sondagem, adequadas para uso em linha, de aplicação quase universal (qualquer molécula contendo C-H, N-H, S-H ou ligações O-H) e, na maioria dos casos, com uma preparação mínima da amostra.⁴

Existem vários modos de aquisição dos espectros em espectroscopia NIR: transmitância, interactância, transfectância e reflectância difusa. A seleção do modo de aquisição ideal depende de diversos fatores, como tipo de amostra e condições do equipamento. A espectroscopia de reflectância mede a luz refletida pela superfície da amostra, que é composta por um componente especular e outro difuso. A reflectância especular, descrita pela lei de Fresnel, contém pouca informação sobre a composição amostral, portanto sua contribuição é minimizada ao se ajustar a posição do detector em relação à amostra. Por outro lado, a reflectância difusa, que é descrita pela teoria de Kubelka-Munk, é a base para a aquisição de espectros por reflectância. O modo de aquisição dos espectros por reflectância difusa é especialmente apropriado para sólidos, e a sua aplicação foi responsável pelo aumento da versatilidade e simplicidade no uso da espectroscopia NIR, contribuindo para a difusão da técnica. Nas medidas de reflectância difusa, o espalhamento e a absorção de luz pelas partículas do sólido contribuem para mudar a intensidade do sinal analítico. Na verdade, a radiação penetra alguns micrometros na amostra, sendo absorvida em função de sua composição química, antes de retornar ao detector. Na grande maioria das aplicações quantitativas envolvendo NIR, ao invés da relação de Kubelka-Munk, usa-se uma função mais prática, $\log(1/R)$, na qual R significa o valor de reflectância. Essa relação é equivalente à absorbância obtida em medidas de transmissão.^{1,4,5}

Desde os primórdios, o uso das informações dos espectros do NIR para efeitos de análise baseou-se na abordagem multivariada para a calibração. Hoje essa interdependência é ainda mais importante, devido ao crescente poder dos microcomputadores e da qualidade dos espectrofotômetros NIR, que permitem que milhares de dados sejam obtidos em questão de segundos. Como a tecnologia NIR propõe uma medida direta da amostra sem pré-tratamento, ela também não apresenta nenhum fator inibitório para aquisição de dados, contribuindo para a geração de grandes conjuntos durante a fase de calibração.⁴

2.4 Espectroscopia no infravermelho médio (MIRS)

A espectroscopia no infravermelho médio monitora as vibrações fundamentais de moléculas, as quais fornecem um perfil químico de um

determinado composto. A região da MIRS, de 4000 a 400 cm^{-1} , é muito robusta e reprodutível, o que permite a medição de pequenas variações dos componentes de uma amostra. As moléculas que absorvem na região do MIR apresentam vibrações em uma ou mais regiões do espectro, dependendo de diversos fatores, como configuração e posição das ligações.⁴

Cada tipo de ligação tem sua própria frequência natural de vibração, e como dois tipos idênticos de ligação em dois diferentes compostos estão em ambientes levemente diferentes, os padrões de absorção no infravermelho médio em duas moléculas de estruturas diferentes nunca são exatamente idênticos. Apesar das frequências absorvidas nos dois casos serem iguais, jamais os espectros infravermelhos serão idênticos.⁵

Um outro uso da MIRS, ainda mais importante, é fornecer informações estruturais de moléculas. As absorções de cada tipo de ligação (N-H, C-H, O-H, C-X, C=O, C-O, C-C, C=C, C \equiv C, C \equiv N, entre outras) são encontradas, geralmente, apenas em regiões específicas do espectro de infravermelho médio⁵. Bandas de absorção características podem ser associadas com componentes majoritários de alimentos. A água é um composto com absorção acentuada na região do MIRS e pode interferir na determinação de outros componentes presentes no alimento. Picos intensos são esperados em 3920, 3490, 3280 e 1645 cm^{-1} , entretanto sua localização exata pode ser influenciada pela presença de solutos, ligações de hidrogênio e pela temperatura. A ligação éster de triglicérides (C-O) ($\sim 1175 \text{ cm}^{-1}$), o grupo C=O ($\sim 1750 \text{ cm}^{-1}$) e a ligação C-H (3000-2800 cm^{-1}) normalmente estão associadas à presença de lipídios, enquanto as bandas de absorção de amidas I ($\sim 1653 \text{ cm}^{-1}$) e II ($\sim 1567 \text{ cm}^{-1}$) têm sido relacionadas com a presença de proteínas e alterações em sua estrutura secundária. Vibrações originadas do estiramento de ligações C-O e C-H na região entre 1100 e 1000 cm^{-1} têm sido utilizadas para identificar açúcares solúveis, enquanto carboidratos mais complexos apresentam bandas de absorção em números de onda mais elevados, como a hemicelulose (1732 e 1240 cm^{-1}), celulose (1170-1150, 1050 e 1030 cm^{-1}), lignina (1590 e 1510 cm^{-1}) e pectina (1680-1600, 1260 e 955 cm^{-1}).⁶

Um fator crítico para a difusão do uso da MIRS foi a sua aplicação na análise de amostras complexas. Historicamente, até o começo dos anos 1980, a aquisição dos espectros esteve restrita as medidas de transmissão em celas de comprimento fixo no caso de amostras líquidas e, para sólidos, à preparação de pastilhas de haletos (KBr, NaCl) ou filmes contendo a amostra. Entretanto, estas técnicas demandavam um considerável preparo e manipulação da amostra, geralmente de maneira destrutiva. A ampliação da aplicabilidade da MIRS veio com o desenvolvimento de novos métodos de aquisição de espectros mais práticos, rápidos e versáteis, como a espectroscopia fotoacústica (PAS), a refletância difusa (DRIFTS) e a reflectância total atenuada (ATR), sendo esta última a mais comumente utilizada.⁶

Com o uso do acessório de ATR, a amostra (sólida ou líquida) é simplesmente colocada em contato com a superfície superior de um cristal de alto índice de refração, normalmente de seleneto de zinco (ZnSe), germânio (Ge), diamante ou iodeto de tálio (KRS-5). O sistema de ATR mede a mudança na intensidade que ocorre em um feixe de radiação infravermelha internamente refletido quando entra em contato com a amostra. Esta interação acontece quando a radiação penetrando no cristal sofre reflexão na sua parte superior, a qual se repete diversas vezes (entre nove e onze vezes ao longo do cristal). Uma onda chamada de onda evanescente é gerada em cada ponto de reflexão. Esta penetra na amostra, interage com a mesma e reduz sua intensidade, produzindo um espectro.⁶

Se comparada com a NIRS, a MIRS possui sinais mais bem resolvidos, mais seletivos e de maior intensidade, o que explica sua aplicação mais antiga e consolidada em análises qualitativas, principalmente de compostos orgânicos (atribuição de bandas a ligações específicas), assim como sua maior sensibilidade, que permite a detecção de componentes em concentrações mais reduzidas. Na análise de amostras menos complexas, é possível inclusive encontrar números de onda seletivos, que permitam desenvolver modelos de calibração univariada, ao contrário do que acontece com o uso da NIRS. Por outro lado, o uso de acessório de ATR diminui a sensibilidade da técnica.^{6,7}

2.5 Análise de imagens digitais

Imagens digitais têm sido utilizadas como fonte de informação analítica desde o século passado. O primeiro trabalho descrevendo a utilização de imagens digitais empregou uma versão rudimentar de um escâner para converter exames médicos em dados digitais, que serviram como base para diagnósticos mais rápidos e acurados.⁸

Mais de vinte anos depois, Geladi e colaboradores publicaram o primeiro artigo que empregou um tratamento quimiométricos de imagens digitais.⁹ Desde então, principalmente nos últimos anos, diversos artigos têm sido publicados utilizando diferentes instrumentos para a aquisição de imagens, tais como celulares, *webcams*, escâners de mesa e câmeras digitais.¹⁰⁻¹⁶

A maneira mais comum de se extrair informações de imagens digitais é a sua decomposição em um sistema de cores, como o RGB. O sistema RGB é um sistema de cores aditivo, que utiliza a combinação das cores Vermelho (*Red*, em inglês), Verde (*Green*, em inglês) e Azul (*Blue*, em inglês), para formar uma ampla variedade de tons (Figura 2).

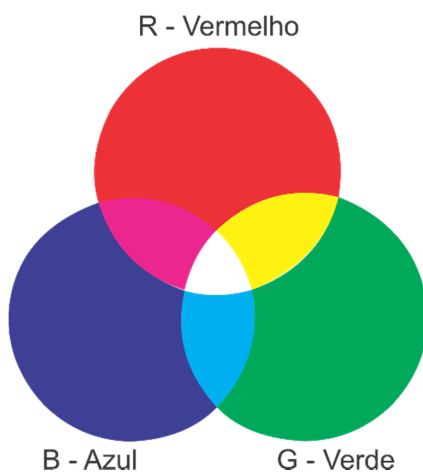


Figura 2 – Representação gráfica do sistema RGB
Fonte: <http://www.jbprint.com.br/blog/imgs/rgb.jpg>

Cada pixel, unidade formadora básica de uma imagem digital, é formado pela combinação dessas três cores. A intensidade de cada cor no sistema RGB é medida em canais (256 para cada cor). O canal 0 significa completa ausência

de uma cor, enquanto o canal 255 representa essa mesma cor em sua intensidade máxima. A combinação dos canais do sistema RGB fornecem 256^3 possíveis tonalidades de tons (Figura 3).

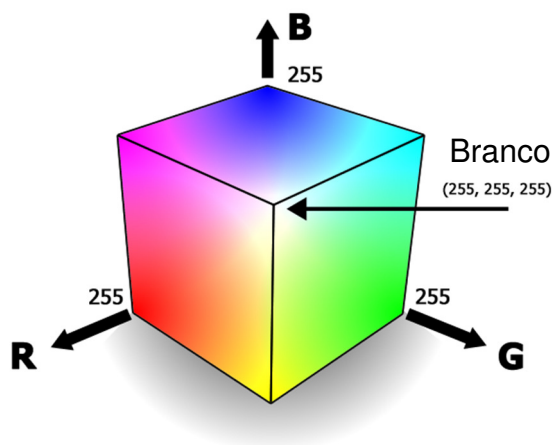


Figura 3 – Formação de tons no sistema RGB

Fonte: <https://fperrotti.wikispaces.com/file/view/cuborgb.jpg/418292514/27x223/cuborgb.jpg>

Após a decomposição de todos os pixels da imagem, a frequência de cada canal é contada, produzindo um histograma de frequências. Este histograma pode ser tratado de maneira similar a dados espectrais utilizados no desenvolvimento de modelos quimiométricos. As variáveis RGB também podem ser fundidas com outros parâmetros de cor, como saturação, intensidade e matiz, resultando em colorgramas.^{12-14,16}

Outra estratégia, que demanda um processamento matemático mais complexo, é a geração de arranjos de dados n-dimensionais (3 dimensões ou mais) que podem ser tratados com métodos de ordem superior.¹⁵ As amostras decompostas no sistema RGB geram arranjos do tipo **OOV** (pixel x pixel x canal), sendo que **O** denota um objeto e **V** um variável. Se as imagens de várias amostras são agrupadas, um arranjo de quarta dimensão é obtido (**OOOV**). Para a simplificação e agilidade no processamento computacional, as variáveis RGB podem ser convertidas para escala de cinzas, reduzindo uma dimensão dos dados, que são então condensados em um arranjo tridimensional (**OOO**) (amostra x pixel x pixel). Um arranjo do tipo **OOO** não é congruente e, por isso, é obviamente não-trilinear. Conseqüentemente, esse tipo de arranjo não pode ser tratado diretamente por métodos de ordem superior, requerendo para isso a

utilização de técnicas de transformação do domínio, como a transformada rápida de Fourier (FFT, *Fast Fourier Transform*) ou a transformada *wavelet*. Estas técnicas convertem arranjos **OOO** em arranjos **OVV** (amostra x componente de frequência x componente de frequência). Esta estratégia normalmente é aplicada quando se quer incorporar ao modelo informações relacionadas com a textura da amostra a ser analisada.^{15,17} Quando as imagens analisadas são homogêneas e não contêm variações texturais, a primeira estratégia descrita, baseada na modelagem direta dos histogramas de frequência de cor, é mais adequada, por ser mais simples e parcimoniosa.

2.6 Quimiometria

De acordo com a definição da ICS (*International Chemometrics Society*), a quimiometria é a disciplina química que emprega métodos matemáticos e estatísticos para planejar ou selecionar experimentos de forma otimizada e para extrair o máximo de informação química a partir da análise dos dados multivariados. O crescimento e desenvolvimento da quimiometria estão intimamente relacionados com o desenvolvimento e a popularização dos microprocessadores, nas décadas de 1970 e 1980, resultando em equipamentos analíticos que permitiram a aquisição de uma grande quantidade de informação sobre uma amostra em pouquíssimo tempo, como, por exemplo, espectros de UV/Visível, de NIR e MID, e até dados mais complexos, como superfícies de fluorescência.^{4,18,19}

De maneira geral, podemos classificar as técnicas quimiométricas em três grandes grupos: planejamentos e otimização de experimentos, calibração multivariada e reconhecimento de padrões, que se pode ser dividido em reconhecimento de padrões supervisionado e não supervisionado.¹⁸

2.6.1 Calibração multivariada

O principal objetivo da calibração multivariada é estabelecer um modelo de regressão relacionando os sinais analíticos medidos para uma amostra (um espectro de NIR, por exemplo), com uma propriedade específica. Os métodos mais consagrados na literatura, todos de calibração inversa, são a regressão linear múltipla (RLM), a regressão de componentes principais (PCR), os

quadrados mínimos parciais (PLS), as redes neurais artificiais (ANN) e as máquinas de suporte de vetores (SVM).

Dentre estes métodos, os mais utilizados são o PCR e o PLS, devido principalmente à sua ampla gama de aplicações e à sua facilidade de uso. Ambos os métodos são capazes de lidar com grande presença de colinearidade e ruídos nos dados, assim como permitem manipular arranjos com mais variáveis que amostras (mais colunas do que linhas na matriz de dados), superando as principais limitações da RLM no tratamento quantitativo de dados espectrais.^{18,20,21}

A principal diferença entre esses dois métodos é a forma pela qual as componentes principais são calculadas. O PLS relaciona a informação contida no bloco **Y**, das variáveis independentes (concentração do analito ou outras propriedades), com o bloco **X**, das variáveis dependentes (espectrais), durante a decomposição dos dados, utilizando os dois simultaneamente. Isto leva a uma pequena perda da ortogonalidade das componentes principais, e com isso, elas recebem a denominação mais abrangente de variáveis latentes. O PCR decompõe **X** separadamente e depois correlaciona seus escores com **Y**, através de uma RLM.^{18,20}

2.6.2 Mínimos quadrados parciais (PLS)

Na construção de modelos PLS, **X** e **Y** são decompostas simultaneamente em uma soma de *h* variáveis latentes. **X** é uma matriz $n \times p$; sendo *n* o número de amostras e *p* o número de variáveis medidas (absorbância em determinados comprimentos de onda, intensidade relativa em determinada razão massa/carga ou valores de diferentes propriedades); **Y** pode ser um vetor (**y**) $n \times 1$, quando o modelo PLS prediz apenas uma propriedade (PLS1), ou $n \times k$, quando o PLS prediz *k* propriedades simultaneamente (PLS2). É importante ressaltar que a denominação PLS1 ou PLS2 não está relacionada com a quantidade de propriedades preditas, e sim com a ordem do tensor dos valores de referência. A base estrutural do modelo PLS é dada pelas seguintes equações.

$$\mathbf{X} = \mathbf{TP}^t + \mathbf{E} = \sum t_h \mathbf{p}_h^t + \mathbf{E} \quad (4)$$

$$\mathbf{Y} = \mathbf{UQ}^t + \mathbf{F} = \sum u_h \mathbf{q}_h + \mathbf{F} \quad (5)$$

Onde \mathbf{T} e \mathbf{U} são as matrizes de escores das matrizes \mathbf{X} e \mathbf{Y} , respectivamente, \mathbf{P} e \mathbf{Q} são as matrizes de pesos (*loadings*) de \mathbf{X} e \mathbf{Y} , respectivamente, e \mathbf{E} e \mathbf{F} são os resíduos de \mathbf{X} e \mathbf{Y} , respectivamente. O coeficiente de regressão linear b correlaciona os blocos \mathbf{X} e \mathbf{Y} de forma linear, de acordo com a equação abaixo:

$$\mathbf{u}_h = b_h \mathbf{t}_h \quad (6)$$

Os valores de b_h são agrupados na matriz diagonal \mathbf{B} , que contém os coeficientes de regressão entre as matrizes de escores \mathbf{T} de \mathbf{X} e \mathbf{U} de \mathbf{Y} . A melhor relação linear possível entre os escores desses dois blocos é obtida através de pequenas rotações das variáveis latentes dos blocos \mathbf{X} e \mathbf{Y} . Os valores preditos para novas amostras (\mathbf{Y}') podem ser então calculados de acordo com a equação 7, com base nos seus escores \mathbf{T}^* .

$$\mathbf{Y}^t = \mathbf{T}^* \mathbf{B} \mathbf{Q} \quad (7)$$

Uma etapa crucial nesse processo é a definição de h (número de variáveis latentes). Normalmente, é utilizada a validação cruzada, um processo de reamostragem, que define o número de variáveis latentes com o qual o modelo apresenta menor erro de predição (RMSECV, raiz do erro quadrático médio de validação cruzada). Os tipos de validação cruzada usados empregam diferentes critérios de reamostragem, de acordo com a natureza dos dados. Para conjuntos de dados pequenos (até em torno de 20 amostras de calibração), usa-se o *leave-one-out*, no qual uma amostra é retirada de cada vez. Já para conjuntos de dados maiores são recomendadas a validação por blocos contínuos, para conjuntos de dados nos quais as amostras estão organizadas em ordem aleatória, por subconjuntos aleatórios, quando as amostras estão organizadas de maneira sistemática, ou por venezianas (*venetian blinds*), que seleciona as amostras retiradas em intervalos constantes, o que é recomendado para conjuntos de amostras organizados em ordem sistemática de agrupamentos/classes. Esta última situação é comum em modelos de classificação supervisionada usando PLS-DA.^{2,18}

2.6.3 N-PLS

O N-PLS (*N-Way Partial Least Squares*, Mínimos Quadrados Parciais Multilineares) é um método de calibração multivariada para tratar dados de ordem superior. O N-PLS pode ser considerado uma extensão do PLS, capaz

de lidar com arranjos de dados de ordem maior do que dois. Ele relaciona um arranjo trilinear de variáveis independentes $\underline{\mathbf{X}}$ (I,J,K) com uma matriz de vetores independentes \mathbf{Y} . Neste método, $\underline{\mathbf{X}}$ é decomposto de maneira trilinear, de forma a encontrar a máxima covariância entre os escores de $\underline{\mathbf{X}}$ e \mathbf{Y} . A principal diferença em relação a outros métodos de segunda ordem, como por exemplo o PARAFAC, é a modelagem simultânea dos dados das variáveis independentes e dependentes.²²

Assim como no modelo PLS, duas etapas são necessárias para a construção do modelo N-PLS, etapa de calibração e etapa de previsão. Na calibração, o arranjo $\underline{\mathbf{X}}$ é decomposto em escores (\mathbf{t}_n) e pesos (\mathbf{w}_j e \mathbf{w}_j), conforme a Figura 4.

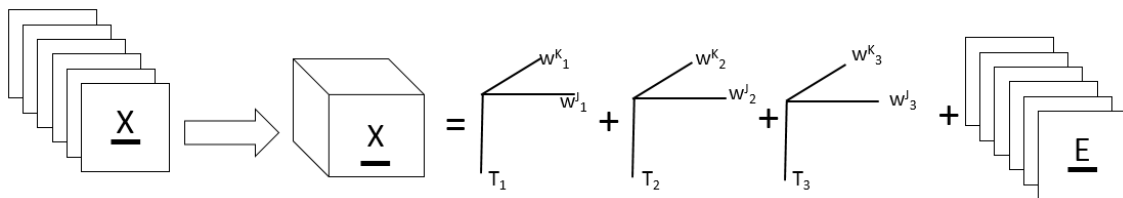


Figura 4 - Decomposição de um arranjo multidimensional através do modelo N-PLS

Onde $\underline{\mathbf{X}}$ é o arranjo tridimensional, $\mathbf{T}(I, F)$ é a matriz relacionada ao modo/dimensão das amostras (escores) e as matrizes $\mathbf{W}^J(J, F)$ e $\mathbf{W}^K(K, F)$ são relacionadas às outras duas dimensões, ou seja, são matrizes de pesos (*weights*) contendo as informações a respeito das variáveis. O arranjo $\underline{\mathbf{E}}$ contém os resíduos, que são a parte não explicada pelo modelo. Esta decomposição está representada na equação 8, nas formas vetorial e matricial.

$$\underline{\mathbf{X}}_{i,j,k} = \sum \mathbf{t}_{ij} \mathbf{w}_{j,f}^i \mathbf{w}_{k,f}^k + \underline{\mathbf{E}}_{i,j,k} = \underline{\mathbf{X}} = \mathbf{T}(\mathbf{W}^K | \otimes | \mathbf{W}^J)^t + \underline{\mathbf{E}} \quad (8)$$

Onde o símbolo “ $| \otimes |$ ” representa o produto de Khatri-Rao, um operador para matrizes de ordem superior²³. A matriz \mathbf{Y} também é decomposta em escores e pesos, representados pela Figura 5 e pela equação 9.

$$\boxed{\mathbf{Y}} = \begin{array}{|c|} \hline \mathbf{q}_1 \\ \hline \mathbf{u}_1 \\ \hline \end{array} + \begin{array}{|c|} \hline \mathbf{q}_2 \\ \hline \mathbf{u}_2 \\ \hline \end{array} + \begin{array}{|c|} \hline \mathbf{q}_3 \\ \hline \mathbf{u}_3 \\ \hline \end{array} + \boxed{\mathbf{F}}$$

Figura 5 - Representação esquemática da decomposição da matriz \mathbf{Y} através do modelo NPLS

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^t + \mathbf{F} \quad (9)$$

Onde \mathbf{Y} é a matriz contendo a propriedade de interesse, \mathbf{U} é a matriz contendo os escores de \mathbf{Y} , \mathbf{Q} é a matriz contendo os *loadings* de \mathbf{Y} e \mathbf{F} é a matriz de resíduos de \mathbf{Y} , que assim como no arranjo $\underline{\mathbf{X}}$, não podem ser explicados pelo modelo.

Para relacionar a matriz de escores de $\underline{\mathbf{X}}$ e a matriz de escores de \mathbf{Y} é necessário encontrar uma relação linear entre a matriz de dados e a matriz da propriedade de interesse. Essa relação é representada pela equação 10.

$$\mathbf{U} = \mathbf{T}\mathbf{B} \quad (10)$$

Onde \mathbf{B} é uma matriz contendo os coeficientes de regressão linear entre o arranjo de dados $\underline{\mathbf{X}}$ e \mathbf{Y} . A matriz $\hat{\mathbf{Y}}$ para as novas amostras pode ser estimada a partir de novos escores \mathbf{T}^* , conforme a equação 11.

$$\hat{\mathbf{Y}} = \mathbf{T}^* \mathbf{b}^t \quad (11)$$

O NPLS apresenta vantagens em relação aos métodos baseados no prévio desdobramento da matriz tridimensional, que são métodos de primeira ordem usados para tratar dados de segunda ordem (*unfold-PLS*, por exemplo). Este tipo de método não é verdadeiramente multidimensional, porque utiliza uma decomposição bilinear para tratar dados multidimensionais. Dentre essas vantagens encontram-se maior robustez, parcimônia, simplicidade do modelo e a facilidade na interpretação dos resultados, além da menor sensibilidade ao ruído.^{15,22,23}

2.6.4 Métodos de classificação supervisionada

Métodos de classificação supervisionada são ferramentas quimiométricas, que buscam desenvolver modelos matemáticos capazes de classificar cada objeto/amostra em sua determinada classe com base em um conjunto de dados, de maneira supervisionada. Uma vez que o modelo de classificação foi gerado, a atribuição de amostras desconhecidas a uma das classes previamente definidas pode ser realizada. Enquanto modelos de calibração fornecem respostas quantitativas, os métodos de classificação são qualitativos. Os métodos de classificação definem uma relação matemática entre variáveis descritivas (variáveis espectrofotométricas, espectros de massa, composição mineral, dentre outras) e uma resposta qualitativa (pertence/não pertence a uma determinada classe).²

Dentre os métodos de classificação supervisionada mais comuns, podemos citar a LDA (*Linear Discriminant Analysis*, Análise discriminante linear), a QDA (*Quadratic Discriminant Analysis*, Análise discriminante quadrática), o PLS-DA (*Partial Least Squares Discriminant Analysis*, Análise discriminante por mínimos quadrados parciais) e o SIMCA (*Soft Independent Modelling of Class Analogies*, Modelos independentes de similaridade utilizando componentes principais).

2.6.5 PLS-DA

O PLS-DA é uma adaptação do método PLS para modelos qualitativos. A principal diferença entre os dois está nas variáveis dependentes, uma vez que no PLS-DA elas representam valores qualitativos (classes) e não quantitativos. Em problemas de classificação binária, a variável Y pode ser arbitrariamente definida como 0 e 1, sendo 1 para amostras pertencentes à classe, e 0 para amostras não pertencentes. O modelo então fornecerá uma resposta numérica, assim como nos métodos quantitativos, e este valor poderá não ser exatamente 1 ou 0. Então, um limite de decisão (*threshold*) precisa ser estabelecido. Este limite pode ser arbitrário (0,5, por exemplo) ou calculado através da teoria bayesiana. O cálculo do limite bayesiano assume que a variância predita de Y segue uma distribuição similar à que será observada nas novas amostras. Amostras com valores preditos abaixo do limite são classificadas como 0 (não pertencente a classe) e amostras com valores preditos acima do limite são classificadas como 1.²⁴

Quando o modelo possuir mais de uma classe, a estratégia acima citada não pode ser utilizada, pois se Y for definido com um número para cada classe (1, 2, 3, 4...N para N classes diferentes, por exemplo), isto significaria uma relação matemática de continuidade entre as classes, como, por exemplo, a classe 2 ser uma classe intermediária entre 1 e 3. A solução é definir um vetor coluna para cada classe e se aplicar o algoritmo PLS1 (cada vetor **y** previsto em separado) ou PLS2 (todos os vetores agrupados em uma matriz) para respostas multivariadas qualitativas. Para cada amostra, o PLS2-DA fornecerá a predição como um vetor de tamanho Nx1, com valores entre 0 e 1. Neste caso, o critério de classificação das amostras é o mesmo para quando só existem duas classes.^{2,24}

2.6.6 Quimiometria aplicada à análise de alimentos

Uma alternativa promissora para superar as desvantagens dos métodos clássicos de análise de alimentos é a utilização conjunta das técnicas acima citadas e quimiometria. Esta combinação pode fornecer métodos que são rápidos, não destrutivos, livres de solventes e reagentes, com baixa ou nenhuma geração de resíduos. Além disso, estes métodos demandam um mínimo de pré-tratamento da amostra e têm suficiente precisão e sensibilidade com menos intervenção humana. Uma restrição à aplicação da NIRS na análise de alimentos é a sua sensibilidade relativamente baixa, pois o limite de detecção desta técnica é, em geral, próximo de 0,1% (p/p)⁴, o que torna inviável a sua utilização na determinação de microcontaminantes, como resíduos de drogas veterinárias, pesticidas, toxinas e outros compostos em baixas concentrações. Situação similar ocorre com outras técnicas mencionadas, como a MIR combinada com ATR e a análise de imagens digitais adquiridas por equipamentos de baixo custo.

Métodos quimiométricos em análise de alimentos podem ser usados, por exemplo, para o controle de qualidade direto (*online*), para separar frutos defeituosos de sadios, na quantificação de diversos parâmetros nas mais variadas matrizes, como leite e derivados,²⁵ frutas e vegetais,^{25,26} azeite,²⁷ mel²⁸ e cervejas.²⁹ Além de quantificação, a quimiometria pode ser útil para detecção de adulterações,³⁰⁻³² determinação de denominação de origem e classificação de diversas matrizes.³³⁻³⁵

2.7 Validação analítica multivariada

O principal entrave para a validação de métodos multivariados tem sido a falta de guias específicos e regulamentação oficial apropriada. Na área de alimentos, tanto os guias oficiais brasileiros³⁶⁻³⁸ quanto os internacionais³⁹⁻⁴² ignoram completamente aspectos multivariados intrínsecos ao desenvolvimento de métodos multivariados. Existe uma necessidade de harmonização de alguns conceitos, como seletividade, sensibilidade e curvas de calibração. Todos os guias tradicionais exigem seletividade total do sinal analítico, o que não faz sentido para calibração multivariada, a qual só tem utilidade na ausência de um sinal seletivo. Recomenda-se também a avaliação da linearidade através das curvas de calibração (resposta analítica *versus* concentração), que não são estritamente possíveis de obter em modelos multivariados. Dentre os esforços feitos nos últimos anos buscando a harmonização de aspectos univariados e multivariados, o conceito mais importante é o NAS (*Net Analyte Signal*, sinal analítico líquido), o qual possibilita separar uma parte específica do sinal total diretamente proporcional à propriedade de interesse, permitindo o cálculo de algumas figuras de mérito (FOM) multivariadas.⁴³

2.7.1 Detecção de amostras anômalas (*outliers*)

Amostras anômalas são amostras que são muito diferentes do restante do conjunto de dados, e sua detecção é crucial no desenvolvimento de modelos multivariados. Estas diferenças podem ser causadas por erros de medição, amostras provenientes de outras populações, etc. A detecção de amostras anômalas pode ser realizada de várias maneiras, e este trabalho adotou uma metodologia robusta para a detecção de amostras com alta influência (*leverage*), altos resíduos no bloco X, isto é, nos dados espectrais, ou altos resíduos no bloco Y, ou seja, os valores de propriedade ou de teor do analito.^{44,45}

Se uma amostra tem um *leverage*, uma medida de sua influência no modelo PLS, maior que um valor limite ($h > 3k/n_c$, onde k é o número de variáveis latentes do modelo PLS e n_c o número de amostras de calibração), ela deve ser considerada uma amostra anômala. Amostras anômalas com altos resíduos em X são detectadas através da comparação do desvio-padrão total, $s(e)$, dos dados espectrais não-modelados com os respectivos desvios padrão de cada amostra,

$s(e_i)$. Se uma amostra apresentar $s(e_i) > 2,5s(e)$, ela deve ser removida com nível de confiança de 99%. Amostras anômalas com altos resíduos em Y podem ser detectadas através da comparação do erro quadrático médio da calibração (RMSEC) com os erros de predição. Se uma amostra possuir um erro de predição maior do que três vezes o RMSEC, ela deve ser removida.^{44,46}

Os dois primeiros testes acima descritos foram aplicados a ambos os conjuntos de calibração e de validação, enquanto que o terceiro teste só foi aplicado ao conjunto de calibração. Amostras anômalas no conjunto de validação também foram detectados pelo teste de resíduos padronizados Jackknife. Este teste estima a variância dos resíduos da regressão independente do ponto avaliado, o que dispensa a necessidade de se calcular várias regressões a cada novo ponto excluído.⁴⁷

2.7.2 Sinal Analítico Líquido (NAS) e curvas pseudounivariadas

O NAS (do inglês *Net Analyte Signal*) é um conceito multivariado proposto por Lorber em 1986⁴⁸, que pode ser definido como a parte do sinal analítico que é exclusivamente relacionada com o analito/propriedade de interesse, e ortogonal ao espaço gerado pelos interferentes (Figura 6).

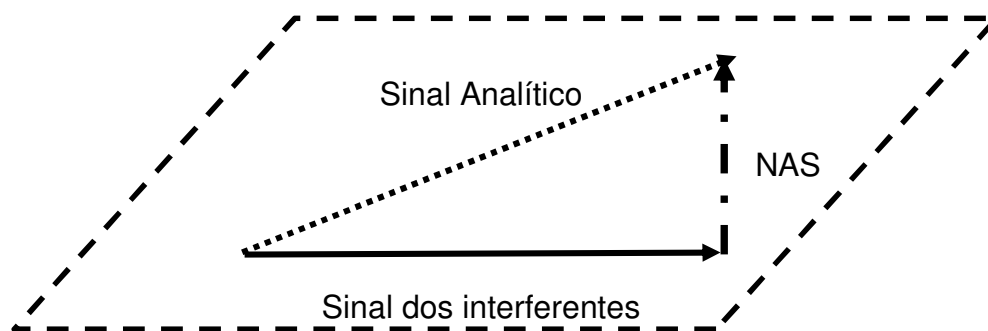


Figura 6 – Representação vetorial do NAS

Inicialmente, este conceito foi proposto para métodos de calibração multivariados diretos, como o CLS (mínimos quadrados clássicos), mas a sua aplicação prática surgiu apenas alguns anos mais tarde, quando esse conceito foi ampliado para os métodos de calibração inversa, tais como PLS e PCR⁴⁹ e melhorado⁴³. O vetor de NAS \hat{x}_{hi}^{nas} pode ser calculado para cada amostra i a

partir dos coeficientes de regressão de um modelo PLS com h variáveis latentes (VL), \mathbf{b} , de acordo com a equação 12.

$$\hat{x}_{hi}^{nas} = \mathbf{b}(\mathbf{b}^T \mathbf{b})^{-1} \mathbf{b}^T \mathbf{x}_i \quad (12)$$

A norma (ou módulo) de cada vetor fornece um $\hat{n}as_i$ escalar para cada amostra, que é análogo a um sinal de analito seletivo univariado.^{42,43,45,49}

O NAS é um conceito útil na validação analítica multivariada porque é usado para calcular algumas Figuras de Mérito (FOM) tais como a sensibilidade, a seletividade e os limites de detecção e quantificação.⁵¹ Este conceito também é útil para traçar as curvas pseudounivariadas de calibração, as quais são uma maneira mais simples de apresentar os resultados obtidos a partir de modelos multivariados, harmonizando-os com o pensamento univariado.⁵²

Para isso, em primeiro lugar, valores escalares de NAS são calculados para as amostras de calibração e, em seguida, um coeficiente de regressão $\hat{\mathbf{b}}_{nas}$ é estimado através de uma regressão linear entre o vetor $\mathbf{n}\hat{\mathbf{a}}\mathbf{s}$, o qual contém as normas dos vetores \hat{x}_{Ai}^{nas} de cada amostra, e o vetor das concentrações do analito ou propriedades (\mathbf{Y}).

$$\mathbf{b}_{nas} = (\mathbf{n}\hat{\mathbf{a}}\mathbf{s}^T \mathbf{n}\hat{\mathbf{a}}\mathbf{s})^{-1} \mathbf{n}\hat{\mathbf{a}}\mathbf{s} \mathbf{b}^T \mathbf{y} \quad (13)$$

Em seguida, o modelo de regressão é expresso como na equação 13, onde \mathbf{e} é um vetor contendo os resíduos.

$$\hat{\mathbf{y}} = \mathbf{b}_{nas} \mathbf{n}\hat{\mathbf{a}}\mathbf{s} + \mathbf{e} \quad (14)$$

2.7.3 Figuras de mérito (FOM)

A validação analítica foi realizada a fim de harmonizar conceitos univariados e multivariados. Neste sentido, este trabalho procurou adaptar os requisitos das diretrizes brasileiras³⁸ e internacionais³⁹⁻⁴² para o estado do arte de como estimar FOM em calibração multivariada.^{44,50} Assim, as seguintes FOM foram estimadas: veracidade, precisão, linearidade, faixa de trabalho, seletividade, sensibilidade, sensibilidade analítica, robustez, viés (*bias*) e relação de desempenho do desvio (RPD, *residual prediction deviation*). Os limites de detecção e quantificação podem ser calculados com base no NAS, mas estes não foram estimados no presente trabalho, nem os conceitos alternativos de limite de decisão ($CC\alpha$) e capacidade de detecção ($CC\beta$), pois eles não são

exigidos para métodos de determinação de analitos que estão presentes em concentrações elevadas.

2.7.3.1 Exatidão

De acordo com o Vocabulário Internacional de Metrologia (VIM), o conceito de exatidão utilizado tradicionalmente foi atualizado. Ele foi substituído pela junção dos conceitos veracidade e precisão. A veracidade é definida como o grau de concordância entre a média de um número infinito de valores medidos repetidamente e o valor de referência. E a somatória da veracidade com a precisão é agora chamada de exatidão, a qual é definida como o grau de concordância entre um valor medido e um valor verdadeiro de um mensurando.⁵³

A veracidade e a precisão podem ser estimadas de forma semelhante para métodos univariados e multivariados. A veracidade foi avaliada através dos erros relativos de predição para cada amostra. No entanto, a veracidade de métodos multivariados é usualmente estabelecida através de parâmetros mais específicos não mencionados nos guias tradicionais de análise de alimentos, tais como RMSEC (raiz do erro quadrático médio de calibração) e RMSEP (raiz do erro quadrático médio de predição). O RMSEP é mais robusto que o RMSEC, uma vez que é estimado a partir de amostras de validação externas ao conjunto de calibração.¹⁸

A precisão dos métodos desenvolvidos nesta tese foi avaliada em dois níveis, repetitividade e precisão intermediária. A repetitividade foi estimada pela análise de três amostras diferentes, com diferentes concentrações, pelo mesmo analista, seis repetições para cada amostra. A precisão intermediária foi estimada por três diferentes analistas que determinaram as mesmas três amostras em seis replicatas, no mesmo dia.⁴⁰

2.7.3.2 Linearidade

Uma vez que não há curvas de calibração para modelos multivariados, a linearidade costuma ser estimada através do coeficiente de correlação (r) entre os valores de referência e os valores preditos. Mas, um valor de r próximo de 1 não é considerado suficiente para assegurar a linearidade do método.⁴⁰ É essencial ainda que se verifique previamente se os resíduos do ajuste apresentam comportamento aleatório.

Esta verificação é frequentemente realizada por inspeção visual de um gráfico dos resíduos, o que pode ser considerado bastante subjetivo. A fim de adotar critérios mais objetivos para avaliar a linearidade, esta tese adaptou a metodologia utilizada por Souza & Junqueira⁴⁷ para o método univariado dos mínimos quadrados ordinários. Assim, o comportamento aleatório dos resíduos do ajuste de predição foi verificado através da aplicação de testes estatísticos, a fim de avaliar a sua normalidade, homocedasticidade, e independência, respectivamente. O principal teste é o de Durbin- Watson (DW),⁵⁴ usado para verificar a ausência de autocorrelação, no qual o desvio padrão dos dados ordenados residuais (SD) é comparado com o desvio padrão das diferenças sucessivas (SD_{sd}) do mesmo conjunto de dados. Se os dados são aleatoriamente e normalmente distribuídos e homoscedásticos, não deve haver diferença estatística entre o SD e o SD_{sd}. Um valor DW é estimado, variando de 0 a 4, com média 2. Se o valor calculado converge para 2, isto significa que não há autocorrelação. Se estes valores afastam-se para 0 ou 4, a autocorrelação aumenta de maneira positiva ou negativa, respectivamente. Os valores de 1,5 e 2,5 podem ser utilizados como os limites de aceitação superiores e inferiores.

A normalidade dos resíduos foi estimada pelo teste de Ryan Joiner (RJ). Para a execução deste teste, os resíduos são ordenados em ordem crescente e plotados contra os valores dos percentis estimados para uma distribuição normal reduzida (quantis normais). Considerando que os resíduos são ordenados para uma amostragem de tamanho n , um gráfico de probabilidade normal, também conhecido como gráfico normal quantil –quantil (ou gráfico Q-Q), é produzido ao se plotar o resíduo de cada amostra (e_i) contra q_i , sendo q_i correspondente aos percentis de uma distribuição normal reduzida. Os quantis normais são obtidos pela equação 15.⁴⁷

$$q_i = \phi^{-1} \left\{ \frac{\left(i - \frac{3}{8}\right)}{\left(n + \frac{1}{4}\right)} \right\}, i = 1, 2, 3 \dots n \quad (15)$$

Onde q_i é o valor normal esperado e ϕ^{-1} é o inverso de uma função de distribuição normal padrão. Então, o coeficiente de correlação entre e_i e q_i é calculado. Coeficientes de correlação abaixo de um valor crítico significam a rejeição da hipótese de normalidade dos resíduos. Coeficientes de correlação

críticos para diferentes níveis de significância podem ser estimados através das equações abaixo:⁴⁷

$$R_{crit} \approx 1,007 - \frac{0,1371}{\sqrt{n}} - \frac{0,3682}{n} + \frac{0,7780}{n^2}, \text{ para } \alpha = 0,10 \quad (16)$$

$$R_{crit} \approx 1,0063 - \frac{0,1288}{\sqrt{n}} - \frac{0,6118}{n} + \frac{1,3505}{n^2}, \text{ para } \alpha = 0,05 \quad (17)$$

$$R_{crit} \approx 0,9963 - \frac{0,0211}{\sqrt{n}} - \frac{1,4106}{n} + \frac{3,1791}{n^2}, \text{ para } \alpha = 0,01 \quad (18)$$

Para a avaliação da homoscedasticidade, foi utilizado o teste de Levene, modificado por Brown e Forsythe (BF). Neste teste, os resíduos de predição são divididos em dois grupos de tamanho n_1 e n_2 . A mediana de cada grupo (\tilde{e}_1 e \tilde{e}_2) é calculada. O valor absoluto das diferenças entre a mediana e cada valor de resíduo é calculado. A média \bar{d}_j e a soma quadrática dos resíduos SQD_j dos d_{ij} valores para cada grupo são calculados. A estatística t_L de Levene é estimada pela equação 19. Quando t_L não excede o valor crítico de $t(1 - \frac{\alpha}{2})$ para $n_1 + n_2 - 2$ graus de liberdade, há indicação de homocedasticidade.⁴⁷

$$t_L = \frac{(\bar{d}_1 - \bar{d}_2)}{\sqrt{(\frac{1}{n_1} + \frac{1}{n_2}) s_p^2}} \quad (19)$$

Onde s_p^2 é a variância combinada, que pode ser estimada pela equação 20.

$$s_p^2 = \frac{(SQD_1 + SQD_2)}{n_1 + n_2 - 2} \quad (20)$$

Como alternativa, os parâmetros da curva de calibração pseudounivariada, introduzida na seção 2.7.2 podem também ser usados para expressar a linearidade.

A faixa de trabalho do método pode ser estabelecida através da confirmação de que o método proporciona graus aceitáveis de veracidade, linearidade e precisão dentro dos limites estudados.

2.7.3.3 Sensibilidade e seletividade do modelo

Para os métodos de calibração multivariada, a seletividade (SEL) é calculada para cada amostra i como a razão entre a norma de seu vetor NAS, $\|\mathbf{nas}_i\|$, e a norma de cada respectivo espectro, $\|\mathbf{x}_i\|$.

$$SEL = \frac{\|\mathbf{nas}_i\|}{\|\mathbf{x}_i\|} \quad (21)$$

À medida que cada amostra tem uma SEL diferente, um valor médio é utilizado para caracterizar o método. Já que métodos multivariados não necessitam de resolução de sinal, também é importante ressaltar que a SEL não é útil para fins de controle de qualidade. Ela apenas quantifica quanto do sinal analítico é utilizado no modelo.⁵⁵

A sensibilidade (SEN) de métodos inversos, tais como PLS e PCR, pode ser calculada como o valor do NAS em concentração unitária, e obtida a partir da norma do vetor dos coeficientes de regressão:

$$SEN = \frac{1}{\|\mathbf{b}\|} \quad (22)$$

Uma vez que a SEN é dependente da técnica analítica empregada, uma FOM mais útil para fins de comparação com outros métodos é a sensibilidade analítica (γ), que é calculada como a razão entre a SEN e o ruído instrumental (ε). No caso das medições de reflectância difusa ou ATR, ε pode ser estimado através do desvio padrão combinado de um certo número de replicatas de espectros da cela vazia, sendo o mesmo válido para as imagens adquiridas por escâner. O ε para as imagens adquiridas pelo celular foi estimado utilizando imagens do suporte sem amostra.⁵⁵

$$\gamma = \frac{SEN}{\varepsilon} \quad (23)$$

O inverso de γ (γ^{-1}) fornece uma estimativa da diferença de concentração mínima que é distinguível pelo método, considerando o ruído aleatório experimental como a única fonte significativa de erro, independentemente da técnica utilizada. Este valor define também o número de algarismos significativos com o qual o resultados do modelo devem ser expressos.⁵⁵

2.7.3.4 Robustez

A robustez avalia a susceptibilidade de um método analítico para pequenas alterações deliberadas em condições experimentais. Para as análises de amostras de queijo utilizando NIRS foram avaliadas os tipos de placas utilizadas na aquisição dos espectros e a forma de processamento da amostra.⁴⁵

2.7.3.5 Viés (Bias)

O viés, ou *bias*, é uma FOM que avalia a presença de erros sistemáticos no modelo, e é calculado apenas para o conjunto de validação, como segue:⁴⁵

$$bias = \sum_{i=1}^{n_v} \frac{(y_i^{ref} - \hat{y}_i)}{n_v} \quad (24)$$

onde y_i^{ref} e \hat{y}_i são os valores de referência e preditos para a propriedade de interesse, respectivamente, e n_v é o número de amostras no conjunto de validação. O desvio padrão dos erros de validação (SDV) também é estimado e utilizado em um teste t de Student para determinar se as estimativas de validação apresentam um viés estatisticamente significativo. O valor t calculado é então comparado com o valor t crítico usando o número apropriado de graus de liberdade, que é igual a n_v .⁵⁵

$$SDV = \sqrt{\frac{\sum [(y_i^{ref} - \hat{y}_i) - bias]^2}{n_v - 1}} \quad (25)$$

2.7.3.6 Relação de desempenho do desvio (RPD)

Outra FOM adequada para avaliar a capacidade de predição dos modelos de calibração é a relação de desempenho do desvio (RDP, do inglês *residual prediction deviation* ou *relative predictive determinant*).⁵⁶ A RPD é a relação da variação natural nas amostras de calibração ou de validação em relação ao tamanho dos prováveis erros que ocorrem durante a predição, e é mais útil para a comparação de modelos em conjuntos de dados diferentes (em diferentes faixas de valores previstos), ou em termos absolutos. Ela é calculada para a calibração e validação, de acordo com as equações 26 e 27, nas quais SD_{cal} e

SD_{val} são os desvios-padrão dos valores de referência para a calibração e validação, respectivamente.^{56,57}

$$RPD_{cal} = \frac{SD_{cal}}{RMSECV} \quad (26)$$

$$RPD_{val} = \frac{SD_{val}}{RMSEP} \quad (27)$$

2.7.3.7 Erro amostral específico de predição (EAEP)

A estimativa da incerteza de predição em calibração multivariada é um assunto muito recente, começando a ser estudada apenas no final da década de 1980.⁵⁸ Desde então, tem havido um interesse crescente em determinar a melhor maneira de estimar a incerteza multivariada. Alguns autores utilizam estratégias de reamostragem, que consistem em simular novos dados matematicamente, a partir dos dados já obtidos, através da adição de resíduos aleatórios (*bootstrap*) ou pela adição de ruído aleatório. Dessa forma, é possível aumentar o número de amostras (normalmente $n > 1000$), e a partir daí, o erro de predição para cada amostra é calculado. Outra abordagem para a estimação do erro de predição é a utilização de equações, como as desenvolvidas por Martens & de Vries em 1995⁵⁹ e Faber & Kowalski.⁶⁰

Em alguma das aplicações desta tese, o EAEP (σ_{PE}) foi calculado com base na equação descrita por Faber & Kowalski.⁵⁸ Esta estratégia permite a estimativa da incerteza de medida para cada amostra, com base em seu *leverage* (h). Esta equação utiliza a incerteza do método de referência para o cálculo da incerteza multivariada, e justamente por isso, escolhemos esta abordagem. A incerteza para cada amostra é calculada de acordo com a equação abaixo, na qual MSEC é o erro médio de calibração, e $V_{\Delta y}$ é a incerteza associada aos valores de referência do conjunto de validação. Quando este valor não está disponível, como é o caso deste trabalho, utiliza-se o RMSECV como estimativa de $V_{\Delta y}$.⁶⁰

$$\sigma_{PE} \approx \sqrt{[(1 + h)MSEC - V_{\Delta y}]} \quad (28)$$

Uma vez que os fatores de um modelo PLS podem consumir mais de um grau de liberdade, o número de pseudo graus de liberdade (PGL) é usado, juntamente com os EPE, na estimativa dos intervalos de confiança. O número de PGL é estimado através da Equação 29, na qual n_c é o número de amostras de calibração.⁶¹

$$PGL = n_c \left(1 - \frac{RMSEC}{RMSECV} \right) \quad (29)$$

2.7.3.8 Intervalos de β -tolerância

Uma ferramenta proposta mais recentemente para a validação de métodos analíticos é a estimativa de perfis de exatidão baseados em intervalos de tolerância de expectativa β (β -TI). Os β -TI foram propostos originalmente para a calibração univariada⁶² e são utilizados como uma ferramenta visual complementar de decisão para avaliar o desempenho de modelos preditivos. Eles garantem que uma fração β (ex. 95%) de todos os resultados futuros obtidos por aquele método apresentará um erro dentro dos limites do intervalo calculado. Então, se o β -TI se encontra contido no intervalo definido como aceitável pelos guias de validação, o método proposto atende os requerimentos necessários para a validação. A utilização do β -TI foi recentemente estendida à calibração multivariada, focando principalmente na validação de métodos utilizando NIRS em análises farmacêuticas.⁶³

O β -TI pode ser estimado através das seguintes equações:

$$\beta - TI = RE(\%)_j \pm t \sqrt{1 + \frac{1}{pn_r B_j^2}} RSD(\%)_j \quad (30)$$

Onde p é o número de séries avaliadas, n_r é o número de replicatas independentes por série, $RE(\%)$ é o erro relativo médio para as n replicatas no j -ésimo nível, $RSD(\%)$ é o desvio padrão relativo para as n replicatas do j -ésimo nível. t é o valor crítico bicaudal de t de Student para v graus de liberdade. v é calculado através da equação abaixo.

$$v = \frac{(R_j + 1)^2}{\frac{(R_j + \frac{1}{n_r})^2}{p - 1} + \frac{1 - \frac{1}{n_r}}{pn_r}} \quad (31)$$

Onde R_j é a razão entre a variância intra-séries e a variância entre series, e B_j é estimado utilizando R_j

$$B_j = \sqrt{\frac{R_j + 1}{n_r R_j + 1}} \quad (32)$$

2.7.4 Validação qualitativa multivariada

Métodos analíticos qualitativos fornecem como respostas variáveis dicotômicas (1/0, Sim/Não, Ausência/Presença, Conforme/Não Conforme). Estes métodos já foram considerados como um ramo da química analítica em declínio nas últimas décadas, principalmente por serem associados de maneira restrita à identificação de íons utilizando reações químicas por via úmida. Porém, com os avanços dos métodos analíticos instrumentais e a necessidade de, muitas vezes, caracterizar a presença ou ausência de um adulterante ou contaminante, independente da sua concentração, métodos qualitativos estão sendo revisitados, sendo identificados como ferramentas promissoras para o desenvolvimento e implementação de métodos de triagem (*screening*) confiáveis⁶⁴. Métodos de triagem simples e rápidos podem confirmar a presença de analitos de interesse, simplificando o processo analítico, e reduzindo tempo e custos, ao reduzir o número de amostras que são submetidas a um processo analítico mais complexo e laborioso de análise quantitativa.^{65,66}

Assim como os métodos quantitativos, métodos qualitativos também devem ser submetidos a procedimentos de validação analítica. Como a resposta fornecida por estes métodos são discretas, testes analíticos e procedimentos diferem daqueles utilizados para a validação de métodos quantitativos. Em 2002, a União Europeia publicou o primeiro guia oficial incluindo a validação de métodos qualitativos.⁶⁷ Desde então, o interesse neste assunto tem crescido continuamente.⁶⁸⁻⁷¹ Com o visível crescimento do número de métodos multivariados de classificação supervisionada, se faz necessária a discussão de como proceder para sua validação, assim como tem sido feito com os métodos de calibração multivariada.⁷² Entretanto, este assunto ainda é insipiente na literatura.

O primeiro passo no processo de desenvolvimento de um método qualitativo é a definição do limite que caracteriza uma amostra como 0

(negativa/não conforme/ausência) ou 1 (positiva/aceita/presença). Métodos qualitativos modernos normalmente se baseiam em leituras instrumentais para a classificação das amostras, e este valor é chamado de limite de corte (LC) (*threshold limit*, em inglês), estimado através da teoria de Bayes.^{69,73} A estimativa do LC foi estendida para métodos multivariados, especificamente para o PLS-DA.⁷⁴ Assumindo que todas as amostras no conjunto de treinamento pertençam a uma das classes definidas, A e B, por exemplo, o seguinte pode ser estabelecido:

$$P(A|y) + P(B|y) = 1 \quad (33)$$

Através da teoria de Bayes, é possível estimar que uma amostra pertence à classe A dado um valor determinado de y , $P(A|y)$, através da seguinte equação:

$$P(A|y) = \frac{P(y|A) * P(A)}{[P(y|A) * P(A) + P(y|B) * P(B)]} \quad (34)$$

Onde $P(A)$ e $P(B)$ são as probabilidades para a observação da classe A ou B no futuro. Se assumirmos que a probabilidade de se observar A ou B é similar a quantas amostras de A ou B estavam no conjunto de treinamento, é possível reduzir a equação acima para:

$$P(A|y) = \frac{P(y|A)}{[P(y|A) + P(y|B)]} \quad (35)$$

e

$$P(B|y) = \frac{P(y|B)}{[P(y|A) + P(y|B)]} \quad (36)$$

Estas duas distribuições tipicamente se interceptam em um único valor, que nos leva a um único ponto no qual $P(B|y)$ e $P(A|y)$ são iguais a 0,5. Este ponto é selecionado como o limite de corte.⁷¹

Após esse limite ser definido, o próximo passo é determinar o número de amostras falso positivas (FP) e falso negativas (FN). FP é o número de amostras que não pertencem à classe e são classificadas como pertencentes, por exemplo, uma amostra que não está adulterada sendo classificada como tal, e FN é o parâmetro oposto. Utilizando FP e FN as seguintes FOM são estimadas:

Taxa de Falso Negativo (TFN), Taxa de Falso Positivo (TFP), Taxa de Seletividade (TSL), Taxa de Especificidade (TES) e Taxa de Confiabilidade (TCF).⁷¹

A TFP é definida como a relação entre FP e a soma de FP com o total de amostras conhecidamente negativas (TN)

$$TFP = \frac{FP}{FP + TN} \quad (37)$$

Analogamente, a TFN é definida como a relação entre FN e a soma de FN mais o número de amostras conhecidamente positivas (TP).

$$TFN = \frac{FN}{FN + TP} \quad (38)$$

A TSL é definida como a relação entre TP e a soma de TP e FN, e a TES como a relação entre TN e a soma de TN e FP. Além disso, a TCF pode ser definida como a diferença entre o total dos resultados (100%) e a soma de TFP e TFN.

$$TSL = \frac{TP}{TP + FN} \times 100 \quad (39)$$

$$TES = \frac{TN}{TN + FP} \times 100 \quad (40)$$

$$TCF = 100 - (TFP + TFN) \quad (41)$$

Apesar de algumas FOM utilizadas nas validações qualitativa e quantitativa terem o mesmo nome, os conceitos e maneiras de avalia-las são diferentes. Enquanto TFP, TFN e TCF expressam a veracidade em análises qualitativas, TSL e TES são consideradas FOM relacionadas com a seletividade do método. Outras duas FOM, Acordância (ACO) e Concordância (CON), expressam a precisão do método. A ACO é a probabilidade de duas amostras idênticas fornecerem a mesma resposta (ambas positivas ou negativas) quando realizadas pelo mesmo laboratório ou batelada, sob condições de repetitividade. Para a sua estimativa, a probabilidade de duas amostras fornecerem a mesma resposta é calculada, e estimada uma média dessa probabilidade para cada batelada, conforme a equação abaixo:⁷¹

$$ACO = \frac{\{k(k-1) + (n-k)(n-k-1)\}}{n(n-1)} \quad (42)$$

Onde n é o número total de amostras em cada batelada e k é o número de resultados positivos em cada batelada. A CON é a FOM qualitativa equivalente à estimativa da precisão intermediária ou reprodutibilidade, em métodos quantitativos. Ela estima a probabilidade de duas amostras de mesma classe fornecerem o mesmo resultado em diferentes laboratórios ou bateladas. A CON pode ser estimada de acordo com a equação abaixo.⁷¹

$$CON = \frac{2[k(k-nb)] + nb(nb-1) - ACO[nb(n-1)]}{n^2b(b-1)} \quad (43)$$

Onde b é o número de laboratórios ou bateladas.^{71,75}

2.7.5 Monitoramento pós-modelo

A grande maioria dos métodos de calibração multivariada publicados na literatura não se preocupa em avaliar o seu desempenho posterior, acompanhando-o por um prazo mais longo. Como muitos trabalhos são puramente acadêmicos, a vida útil desses modelos não costuma ser avaliada. No entanto, em aplicações “reais” uma das primeiras preocupações práticas dos responsáveis pelo laboratório de controle de qualidade é a questão: “Até quando o modelo vai continuar funcionando?” Nesse sentido, esta tese se propôs a fazer um monitoramento pós-modelo em sua primeira aplicação, a qual foi desenvolvida em situação real, numa parceria com um laboratório de controle fiscal do Governo de Minas Gerais.

As amostras utilizadas para construir os modelos foram obtidas a partir de 2009. Após os métodos baseados em NIRS terem sido estabelecidos e validados, eles foram então implementados em substituição aos métodos tradicionais no IMA, a partir de 2011. O desempenho dos modelos foi monitorado por amostragem aleatória de, no mínimo, uma amostra por semana, as quais também foram determinadas pelos métodos de referência. A veracidade destas determinações foi avaliada de acordo com as normas brasileiras.³⁶

As cartas de controle foram contruídas com base nos erros relativos de predição. De acordo com essa normatização, dois terços das amostras podem apresentar erros relativos dentro da faixa de $\pm 15\%$ (limites de alerta), enquanto um terço pode ter erros máximos admissíveis abaixo de $\pm 33\%$ (limites de ação). Caso o limite seja ultrapassado, o modelo deve ser revisado.³⁶

3. Modelos para o controle fiscal de queijos utilizando espectroscopia NIR e PLS

A grande maioria dos modelos multivariados propostos para a análise de alimentos não são suficientemente robustos para uso em análise de rotina, necessitando incorporar mais variância e ter sua confiabilidade testada em condições reais. Poucos trabalhos publicados apresentaram um método multivariado que efetivamente fosse usado e testado no dia a dia de um

laboratório de rotina, ou mesmo realizaram um processo de validação analítica, garantindo sua confiabilidade.

Neste capítulo, duas aplicações foram desenvolvidas para o controle de parâmetros físico-químicos de qualidade de queijos: i) determinação de umidade e gordura em queijo muçarela e ii) determinação de umidade em queijo minas. No desenvolvimento destes métodos utilizaram-se amostras oriundas da rotina analítica do IMA, visando justamente incorporar toda variabilidade presente em uma situação real de um laboratório de controle fiscal. Essa amostragem incluiu diferentes produtores, diferentes receitas e a própria sazonalidade do produto, uma vez que as amostras foram coletadas ao longo de dois anos, com isso incorporando a variabilidade inclusive ocasionada pelas épocas de chuva e de seca, que alteram as características nutricionais do leite.

3.1 Determinação de umidade e gordura em queijos muçarela

A Muçarela é um queijo originalmente feito de leite de búfala pasteurizado e parcialmente desnatado, amplamente encontrada em todo o mundo, e originário da região de Campania, no sul da Itália. É o queijo mais consumido nos EUA, Brasil e outros países, principalmente devido ao seu uso em pizzas, sanduíches e lanches.⁷⁶

No Brasil, o Estado de Minas Gerais é o maior produtor de queijo, com cerca de metade da produção (215 mil toneladas por ano, em 2001).⁷⁷ Desse total, 24% corresponde a muçarela. A produção de muçarela no Brasil é regulada pelas diretrizes do Ministério da Agricultura, Pecuária e Abastecimento (MAPA), que estabelecem limites para a gordura (mínimo de 35% p/p, em base seca) e umidade (máximo de 60% p/p).⁷⁸

De acordo com esta legislação, o método de referência para a determinação de gordura é o de Rose Gottlieb, que consiste em extrair os compostos lipídicos do leite através de sucessivas extrações com solventes orgânicos de diferentes polaridades, seguidas de evaporação dos solventes e pesagem dos compostos lipídicos residuais. Por ser um método demorado, muitas vezes ele é substituído pelo método de Gerber em rotinas laboratoriais.⁷⁹ Este método baseia-se na utilização de um butirômetro (Figura 7), e possui desvantagens. A determinação não pode ser automatizada e envolve um certo

risco na manipulação de ácido sulfúrico concentrado⁷⁸. Além disso, a manipulação do butirômetro requer prática, o que tem efeito negativo sobre a robustez e a precisão do método.⁸⁰

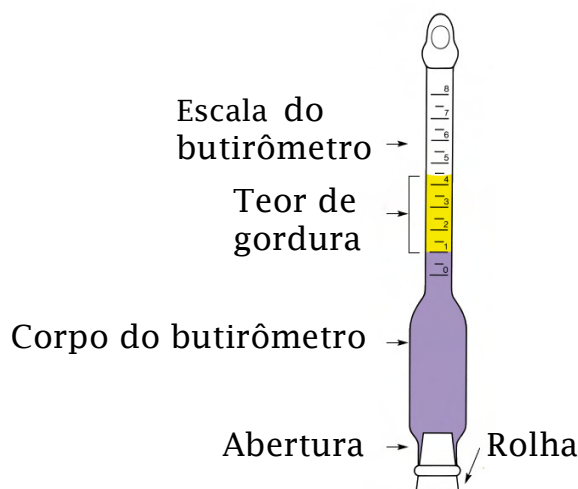


Figura 7 – Butirômetro de Gerber

A determinação de umidade é realizada pelo método gravimétrico de dessecação em estufa⁷⁸. Embora este método seja simples, a sua principal desvantagem é a lentidão, uma vez que cada análise demora entre 3 e 6 horas.⁸¹

Em geral, os métodos analíticos para a determinação de macroconstituintes, tais como gordura e umidade, são lentos, relativamente caros, e exigem pessoal altamente qualificado. Estes métodos clássicos não são eficazes o suficiente para cobrir a crescente demanda e os baixos custos necessários no controle de qualidade de alimentos em indústrias ou na inspeção de qualidade pelos órgãos públicos.⁸²

O primeiro trabalho a sugerir um método baseado em NIRS para a quantificação em produtos lácteos foi publicado por Ben Gera & Norris, em 1968,⁸³ os quais estimaram o conteúdo de gordura no leite por meio de regressão linear múltipla (RLM). Mas, em geral, a utilização mais ampla de métodos de NIRS só surgiu no final dos anos oitenta, quando modelos de calibração multivariada mais estáveis e robustos, como o PLS e o PCR, substituíram a RLM na maioria das aplicações práticas. Em 1994, Pierce & Wehling⁸⁴ publicaram o primeiro trabalho desenvolvendo um método usando NIRS e PLS para determinar umidade e gordura em queijo cheddar. Desde então, vários artigos foram publicados usando NIRS e calibração multivariada para predição destes

parâmetros de qualidade, tais como gordura, proteína, umidade e outros, em diferentes tipos de queijos (fresco, Danbo, Emmental e Cheddar) e derivados.^{7,85-90} Especificamente, nenhum artigo foi publicado desenvolvendo métodos baseados em NIRS para a determinação de parâmetros em queijos muçarela.

Muitos modelos de calibração multivariada baseados em NIRS não são confiáveis, robustos e suficientemente estáveis para serem usados na prática em análises de rotina. A maioria dos artigos citados acima não incorporou variância suficiente em seus modelos, necessária para um uso mais amplo e robusto na rotina do controle de qualidade fiscal. Daí, a necessidade de desenvolver modelos mais robustos, para os quais a estratégia de incorporar amostras com uma grande diversidade de origem é fundamental.⁹¹. Nesses trabalhos, os modelos foram desenvolvidos a partir de amostras de queijo obtidas de um pequeno número de produtores (às vezes, de um único produtor). Absolutamente nenhum dos modelos desenvolvidos nesses trabalhos foram aplicados em situações reais de análises de rotina ou tiveram sua robustez avaliada nesse contexto. Além disso, uma validação analítica multivariada não foi realizada em nenhum dos artigos previamente citados. Esta deficiência é encontrada não apenas em trabalhos relacionados a área de bromatologia, mas na grande maioria dos artigos envolvendo calibração multivariada. Talvez, a falta dessas avaliações mais rigorosas seja um dos fatores que contribui para a não utilização destas técnicas em situações de rotina, e também para falta de reconhecimento de métodos multivariados pelos órgãos oficiais, que, salvo raras exceções, negligenciam por completo a calibração multivariada.

3.1.2 Objetivos

Desenvolver modelos PLS robustos para a determinação de gordura e umidade em queijos muçarela utilizando NIR e PLS, e validá-los de acordo com a legislação vigente, tentando harmonizar o guia do MAPA com a realidade multivariada.

3.1.3 Materiais e métodos

3.1.3.1 Equipamentos e softwares

Os espectros de infravermelho foram obtidos em um espectrofotômetro Perkin Elmer FTIR 400 (Waltham, EUA), equipado com um acessório de reflectância difusa. Os dados foram tratados usando o *software* MATLAB, versão 7.13 (MathWorks, Natick, EUA) juntamente com o pacote PLS Toolbox versão 6.5 (Eigenvector Technologies, Manson, USA). Para a detecção de amostras anômalas utilizou-se uma rotina desenvolvida por Jez W. B. Braga (IQ-UnB).⁹²

3.1.3.2 Materiais, reagentes e amostras

Todos os reagentes e materiais (ácido sulfúrico, álcool isoamílico e areia purificada) foram de grau analítico, adquiridos de fornecedores certificados e utilizados sem purificação adicional. Um total de 123 amostras de queijo muçarela, fabricados por diferentes processos e originárias de mais de 100 diferentes produtores de todo o Estado de Minas Gerais foram coletadas e analisadas entre 2009 e 2010. As amostras de queijo foram transportadas sob refrigeração a 4 °C e analisadas imediatamente após a sua abertura no laboratório. Antes da abertura, as amostras eram deixadas em repouso para atingir o equilíbrio térmico por um período de 30 minutos. As fatias da amostra foram retiradas da parte interna do queijo, evitando a parte externa, que está sujeita a ressecamento e degradação.

3.1.3.3 Procedimentos

Todos os procedimentos foram realizados no Laboratório de Análises Físico-Químicas (LAFQ) do Instituto Mineiro de Agropecuária (IMA), localizado em Contagem/MG. As amostras de queijo foram cortadas, trituradas com um misturador industrial, e transferidas para uma placa Perkin Elmer®, em quantidade suficiente para cobrir seu fundo. Os espectros foram coletados como uma média de 32 varreduras, de 1000 a 2500 nm, passo 2 nm, com resolução de 8 cm⁻¹. Dez espectros idênticos da placa vazia foram também registradas nas mesmas condições, a fim de estimar o ruído instrumental. Durante todas estas medições, a temperatura da sala foi controlada a 22 ± 1 °C.

As análises por métodos de referência foram realizados de acordo com a literatura,⁷⁹ sem repetições. A determinação de gordura foi realizada com o método de Gerber. Um grama de uma amostra de queijo foi pesada com precisão no butirômetro, no qual 10 ml de uma solução de ácido sulfúrico ($d=1,605 \text{ g.cm}^{-1}$) e 1 mL de álcool isoamílico foram adicionados, então, a vidraria foi fechada, agitada e aquecida durante 15 minutos num banho de água a 65 °C; a agitação e o aquecimento foram repetidos até digestão total da amostra pelo ácido e, finalmente, a fase lipídica foi ajustada para a leitura no ponto zero da vidraria.

A determinação da umidade foi realizada por método gravimétrico (dessecação em estufa). Três gramas de uma amostra de queijo foram pesadas numa placa e misturadas com a mesma quantidade de areia fina purificada pré-seca. A adição de areia auxilia na secagem mais rápida e eficiente da amostra, evitando a formação de crostas que dificultam a evaporação da água. Esta mistura foi homogeneizada e seca em estufa a 105 °C, até peso constante (variação menor que $\pm 0,001 \text{ g}$), o que costuma durar de três a seis horas.

3.1.4 Resultados e discussão

3.1.4.1 Desenvolvimento do modelo PLS

Os espectros de infravermelho próximo das 123 amostras de queijos muçarela analisadas são mostrados na Figura 8. Ao observar estes espectros, é possível atribuir as bandas mais intensas e largas centradas em 1454 e 1936 nm às vibrações das moléculas de água, que correspondem ao primeiro sobretom de estiramento da ligação OH, e à combinação das bandas vibracionais dos modos de estiramento e deformação angular de ligações O-H respectivamente.^{83,84}

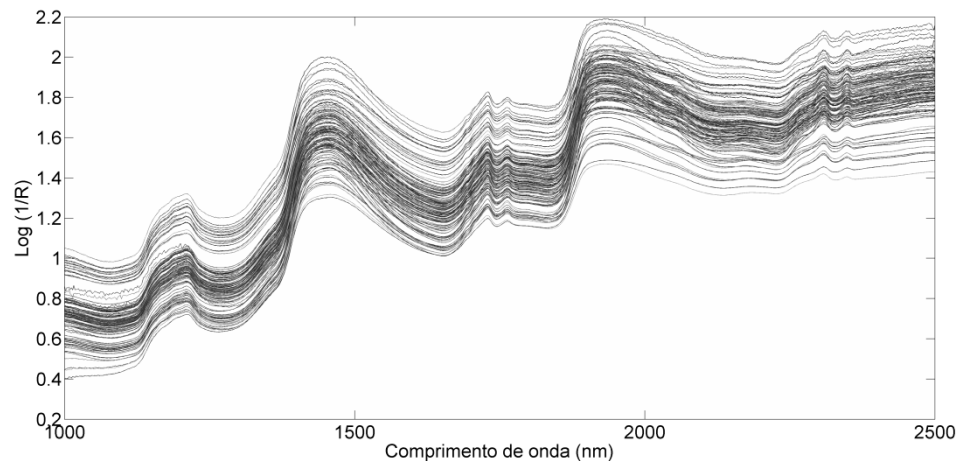


Figura 8 – Espectros das 123 amostras de queijo muçarela

Outros picos observados podem ser relacionados com o teor de gordura dos queijos e a vibrações das moléculas lipídicas, tais como: a banda centrada em 1210 nm, que é atribuída ao segundo sobretom do estiramento de ligações C-H; os picos em 1730 e 1763 nm, relacionados com o primeiro sobretom de estiramento C-H, e os picos em 2308 e 2348 nm, atribuídos a bandas de combinação resultantes de estiramento C-H e deformação de um grupo CH_2 .^{83,84}

As amostras foram divididas em 82 para o conjunto de calibração e 41 para o conjunto de validação usando o algoritmo de Kennard-Stone,⁹³ que garante a presença das amostras mais representativas no conjunto de calibração. O algoritmo define, inicialmente, as duas amostras mais distantes entre si, utilizando a distância euclidiana. Para selecionar novas amostras para o conjunto, o algoritmo seleciona a amostra mais distante daquelas previamente selecionadas. Este processo é repetido até que o número de amostras, definido previamente, seja atingido.

Na Figura 8 também pode ser observada a presença de desvios não-lineares de linha de base, típicos em medidas de reflectância difusa, causados pelo espalhamento multiplicativo da luz. Assim, os dados foram pré-processados usando correção de espalhamento multiplicativo (MSC), a fim de superar este problema,⁹⁴ antes de serem centrados na média. O espectros pré-processados por MSC são mostrados na Figura 9.

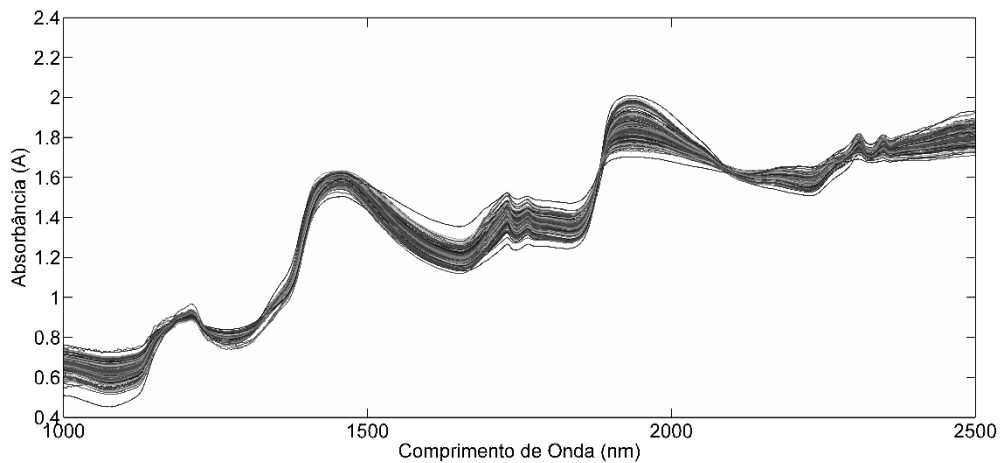


Figura 9 - Espectros de queijo muçarela pré-processados com MSC

O número de VLs em cada modelo foi escolhido por validação cruzada usando divisão em blocos contínuos (9 divisões), com base nos valores mais baixos de RMSECV (raiz do erro quadrático médio de validação cruzada). Os melhores modelos foram inicialmente selecionados, com 4 e 7 VLs para umidade e gordura, respectivamente.

Modelos PLS por intervalos (iPLS com 10 intervalos) também foram testados, mas todos os resultados foram piores (em geral, maiores valores de RMSECV e maior número de VLs) do que os modelos PLS utilizando toda a faixa espectral. Os coeficientes de regressão dos modelos PLS para determinação de gordura e umidade são mostrados nas Figuras 10 e 11, respectivamente. Ao observar a Figura 10, é possível identificar as regiões espectrais que mais contribuíram para a predição de gordura. Como era de se esperar, os maiores coeficientes foram relacionados aos segundo e primeiro sobretons de estiramento de ligações C-H lipídicas (cerca de 1210, 1730 e 1763 nm); no entanto, é interessante notar que a contribuição mais intensa veio da região entre 1870 e 1890 nm, que pode ser atribuída à banda de combinação de vibrações de C-O e O-H de ácidos carboxílicos,⁹⁵ os quais estão presentes na muçarela em baixos teores.

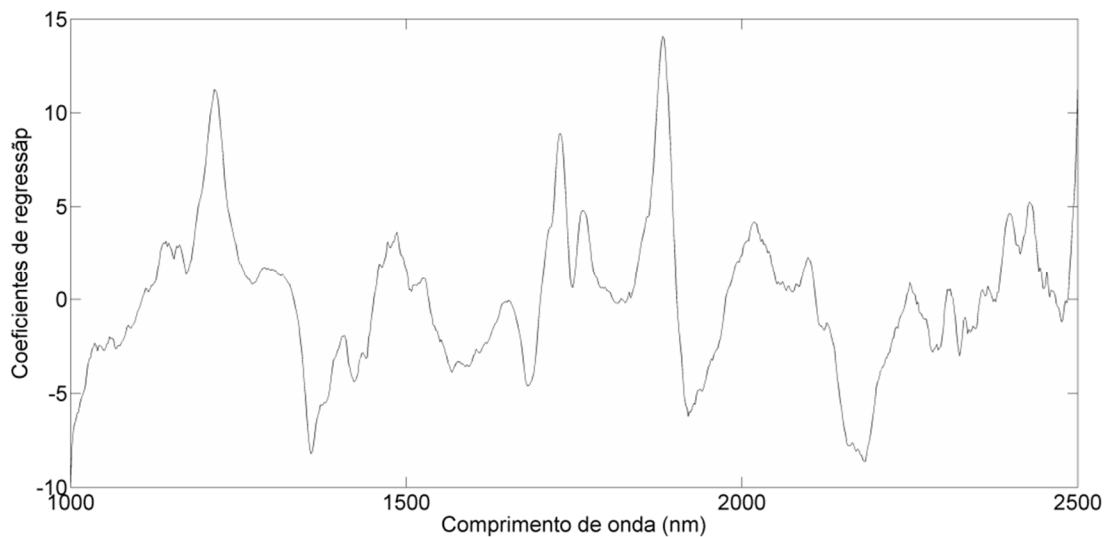


Figura 10 – Coeficientes de regressão para o modelo de determinação de gordura em queijo muçarela

Na Figura 11 podem ser vistas as regiões que mais contribuem para a predição da umidade. Os coeficientes são mais positivos entre 1850 e 1910 nm, correspondendo à banda de combinação de O-H de água; por outro lado, os coeficientes mais negativos foram relacionados às vibrações C-H lipídicas, uma vez que existe uma relação inversa entre água e gordura na composição de queijos. Essas vibrações foram observadas em torno de 1440 nm, e entre 1690 e nm 1780, atribuídas à combinação de primeiro sobretom de C-H, e ao primeiro sobretom de estiramento C-H, respectivamente.

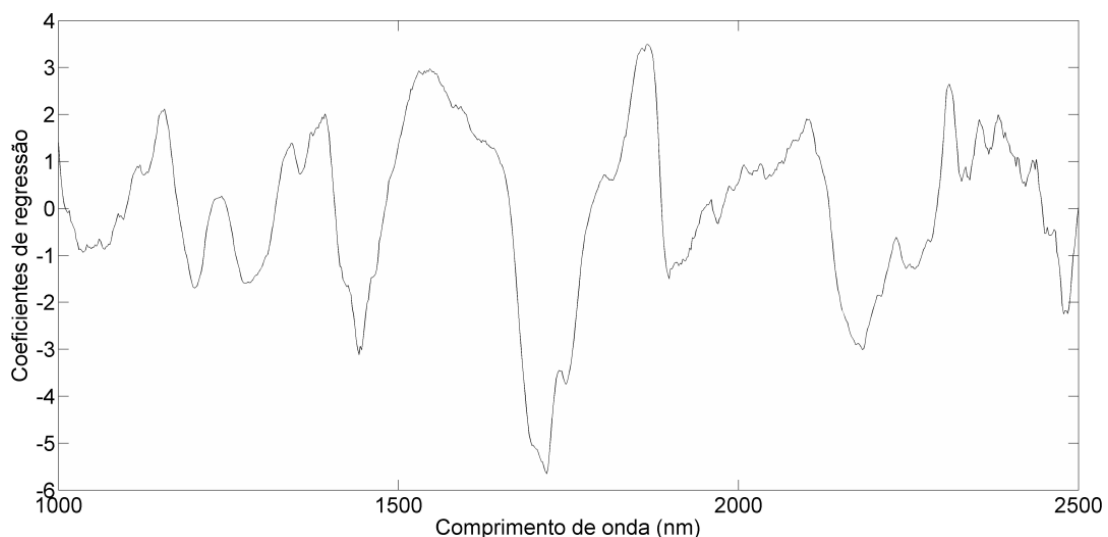


Figura 11 - Coeficientes de regressão para o modelo de determinação de umidade em queijo muçarela

3.1.4.2 Otimização dos modelos pela detecção de amostras anômalas

Após a construção dos modelos PLS, foi realizada a otimização utilizando o procedimento para a detecção de amostras anômalas descrito previamente. Os valores extremos foram detectados a um nível de confiança de 99%, e os resultados foram resumidos na Tabela 1. As amostras anômalas podem ser removidas dentro de um limite de 2/9 (22,2 %) de seu número total, conforme estabelecido pelas diretrizes brasileiras.³⁶

Tabela 1 - Otimização dos modelos PLS pela detecção de amostras anômalas (modelos finais estão destacados em negrito)

Modelos	1	2	3	4
Gordura				
Amostras de calibração	82	77	74	---
Amostras de validação	41	41	37	---
Variáveis latentes	7	8	8	---
RMSEC (%)	3,1	2,1	1,8	---
RMSEP (%)	3,2	3,0	2,1	---
Umidade				
Amostras de calibração	82	77	74	73
Amostras de validação	41	41	41	35
Variáveis latentes	4	8	8	6
RMSEC (%)	1,8	1,0	0,8	0,6
RMSEP (%)	2,2	1,8	1,7	0,9

A otimização do conjunto de validação foi realizada após a otimização do conjunto de calibração. Além disso, não mais de três rodadas de detecção de amostras anômalas (quatro modelos) devem ser realizadas, a fim de evitar o "efeito bola de neve" (*snowballing*), que ocorre quando são sempre identificados novos *outliers* a cada novo modelo.⁴⁴

Como pode ser visto na Tabela 1, para o modelo de gordura, oito *outliers* foram detectados no conjunto de calibração (correspondente a 9,8% das amostras) e quatro no conjunto de validação (também 9,8%); destes 12 *outliers*, três foram detectados com base no alto *leverage* e nove com base nos resíduos em Y. Para o modelo de umidade, nove *outliers* foram identificados no conjunto de calibração (11,0%) e seis no conjunto de validação (14,6%). Destes quinze

outliers, três foram detectados com base no *leverage*, um com base nos resíduos espectrais (X) e onze com base nos resíduos em Y.

Portanto, o modelo PLS otimizado para a determinação de gordura foi construído com 74 amostras de calibração e 37 amostras de validação, usando 8 VLs, explicando 99,1% da variância em X e 89,8% em Y. O modelo otimizado para determinação de umidade foi construído com 73 amostras de calibração e 35 amostras de validação, utilizando 6 VLs, explicando 98,2% da variância em X e 94,1% em Y.

3.1.4.3 Validação analítica

A Tabela 2 resume os parâmetros estimados para a avaliação das FOM para os métodos propostos. Nas subseções seguintes, cada FOM estimada é discutida especificamente.

3.1.4.3.1 Linearidade

Tal como discutido na subseção 2.7.3.2, a linearidade foi estimada através do ajuste dos valores de referência contra os valores preditos. Os resíduos para os modelos de gordura e de umidade estão apresentados nas Figuras 12 e 13. Na maioria da literatura referente à calibração multivariada, os resíduos são plotados em função dos valores de referência, o que não é estritamente correto. Os resíduos devem ser representados graficamente em função dos valores preditos, uma vez que não existe correlação entre eles ($r=0$). Se os resíduos são plotados contra os valores de referência, sempre haverá correlação, com o coeficiente de correlação (r) sendo igual a um menos o valor de r da curva de calibração.⁹⁶

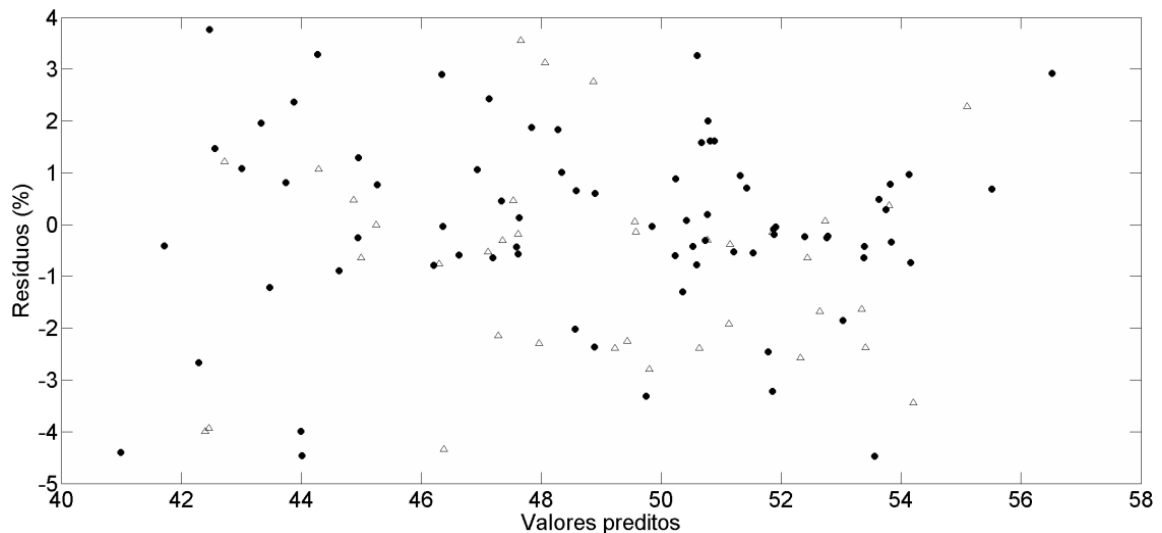


Figura 12 – Gráfico de resíduos vs. valores preditos para o modelo de determinação de gordura – amostras de calibração (●) e de validação (Δ)

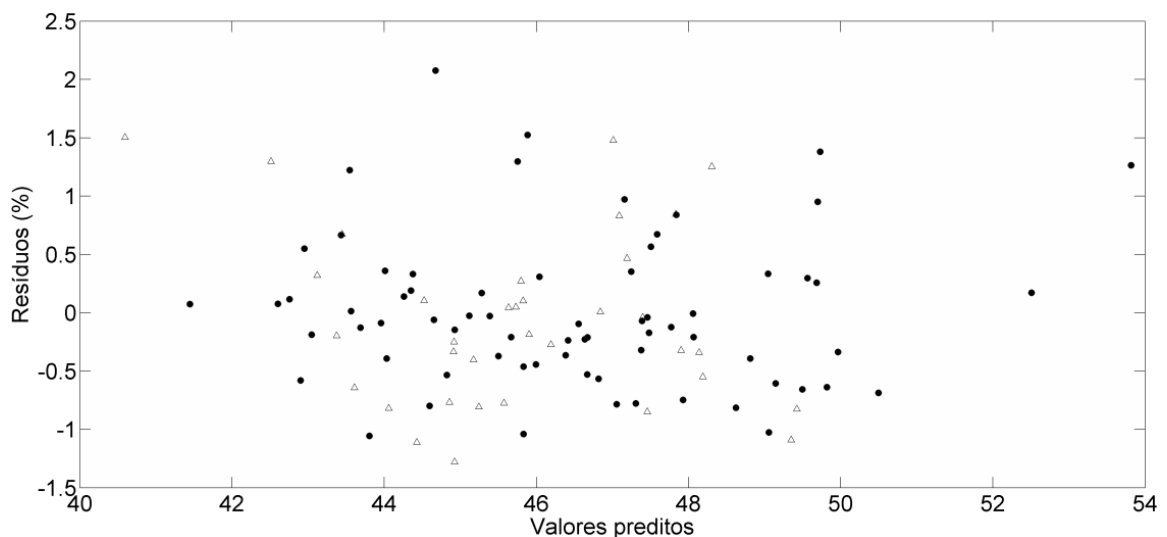


Figura 13 - Gráfico de resíduos vs. valores preditos para o modelo de determinação de umidade – amostras de calibração (●) e de validação (Δ)

Como pode ser observado nas figuras acima, nenhuma tendência sistemática é evidente nas distribuições residuais. No entanto, o comportamento aleatório destes resíduos foi confirmada por testes estatísticos apropriados.⁴⁷ No nível de confiança de 95%, os resíduos dos dois modelos foram considerados normais e homoscedásticos, de acordo com os testes RJ e BF, respectivamente. Além disso, a independência dos resíduos foi verificada pelos valores do teste DW (Tabela 2), os quais estavam dentro dos limites da faixa de aceitação, entre 1,50 e 2,50.

Tabela 2 - Figuras de mérito, estimadas para os modelos PLS de determinação de gordura e umidade

Figuras de Mérito	Parâmetros	Gordura	Umidade
		Valores	
Veracidade (n=40)	Erro relativo médio	2,1 %	0,1 %
Precisão (n=18)	DPR repetitividade ^a (DPR _r)	2,6 % (1,8 %)	1,4 % (1,8 %)
		3,2 % (1,7 %)	1,8 % (1,8%)
		3,6 % (1,8 %)	1,7 % (1,7 %)
	DPR precisão intermediária ^b (DPR _R)	3,1 % (2,7 %)	1,6 % (2,6 %)
		8,5 % (2,6 %)	2,2 % (2,8 %)
		4,1 % (2,7 %)	1,7 % (2,5 %)
Linearidade	Durbin-Watson	1,8	1,9
	Inclinação	1,00 ± 0,02	1,00 ± 0,03
	Intercepto	0,01 ± 2,70	0,00 ± 1,40
	Coefficiente de correlação (r) ^c	0,9442	0,9702
Faixa de trabalho		38,7 a 58,0 %	41,5 a 55,1 %
Seletividade do modelo		0,07	0,11
Sensibilidade do modelo		0,009 ^d	0,012 ^d
Sensibilidade analítica (γ)		1,8 % ⁻¹	2,4 % ⁻¹
γ ⁻¹		0,60%	0,40%
Viés		-0,775±1,942 %	0,001±0,933 %
RPD	RPD calibração	2,1	2,9
	RPD validação	2,0	2,4

^a Resultados para três amostras em três diferentes concentrações analisadas no mesmo dia por um mesmo analista, em sextuplicata. Valores limites de DPR_r (entre parênteses) calculados através da equação de Horwitz para cada nível de concentração estudado.

^b Resultados para três amostras em três diferentes concentrações em dias diferentes por analistas diferentes, em sextuplicata. Valores limites de DPR_R (entre parênteses) calculados através da equação de Horwitz para cada nível de concentração estudado.

^c Parâmetros do ajuste para a curva de valores de referência vs. valores preditos para as amostras de calibração

^d Valores expressos como a razão entre log (1/R) e %

Uma vez que os modelos lineares foram considerados válidos, através da avaliação dos testes estatísticos, os parâmetros do ajuste da curva de valores de referência contra os valores previstos puderam ser calculados (Tabela 2). Outra forma de expressar a linearidade de modelos multivariados NIRS é através das curvas de calibração pseudo-univariada, que são mostradas para os modelos de gordura e de umidade nas Figuras 14 e 15. Os ajustes das curvas são mostrados nas equações 19 e 20, respectivamente.

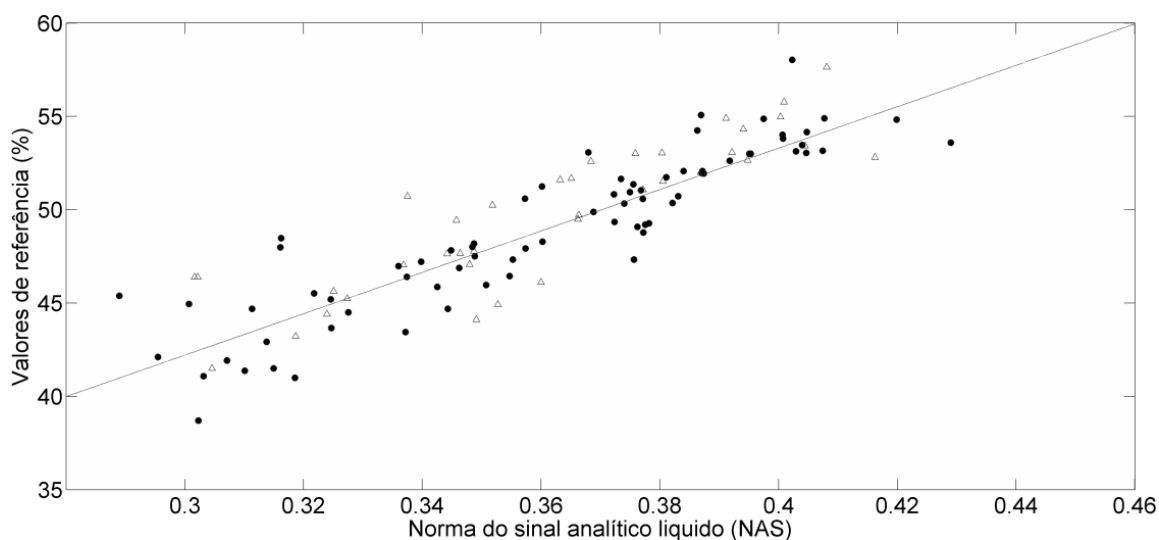


Figura 14 – Curva de valores de referência vs. norma do NAS para o modelo de determinação de gordura - amostras de calibração (●) e de validação (Δ)

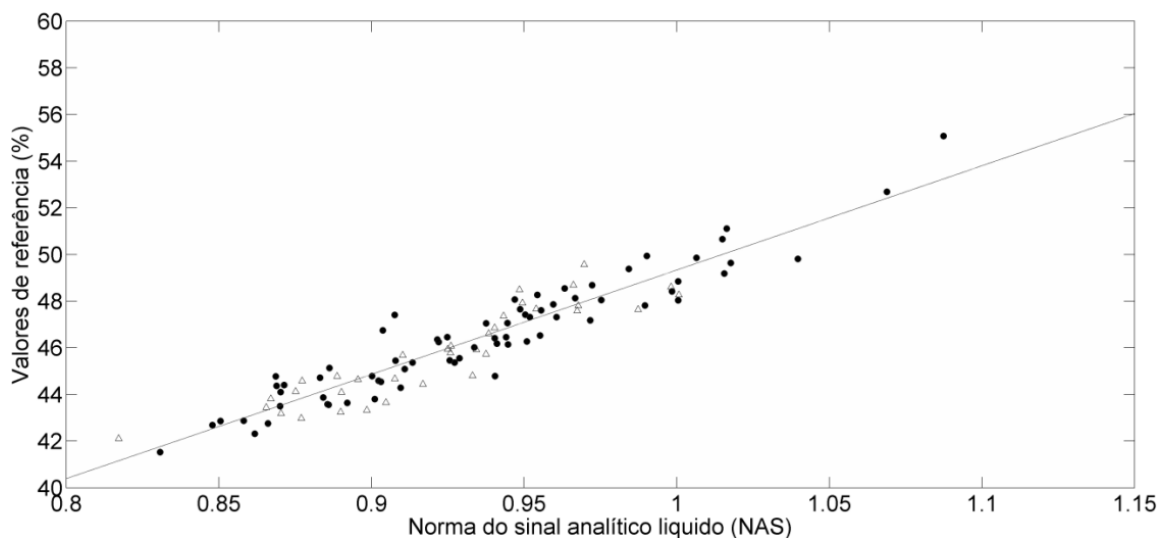


Figura 15 - Curva de valores de referência vs. norma do NAS para o modelo de determinação de umidade - amostras de calibração (●) e de validação (Δ)

Este tipo de representação permite uma melhor compreensão dos modelos PLS, de uma forma univariada, o que é útil em análises de rotina. Os valores de norma do NAS são equivalentes à intensidade de sinais analíticos seletivos obtidos, após a eliminação da contribuição dos interferentes, a partir do vetor do espectro de cada amostra.

$$[Gordura] = 110,91|n\hat{a}s| + 8,94 \quad (44)$$

$$[Umidade] = 47,69|n\hat{a}s| + 4,45 \quad (45)$$

3.1.4.3.2 Veracidade, precisão e faixa de trabalho

A veracidade dos métodos pode ser avaliada através dos valores de RMSEP (absolutos) 2,1 e 0,9%, para gordura e umidade, respectivamente (Tabela 2). Ela também pode ser avaliada com base nos erros relativos médios observados, 2,1 % (variando entre -8,7 e 8,1%) para a determinação de gordura, e 0,1% (variando entre -2,9 e 3,5%) para a determinação de umidade (Figuras 16 e 17). Os resultados encontrados estão de acordo com o esperados, pois para a faixa de concentração do analito, os erros relativos médios podem variar entre -3 e 3%.⁴¹

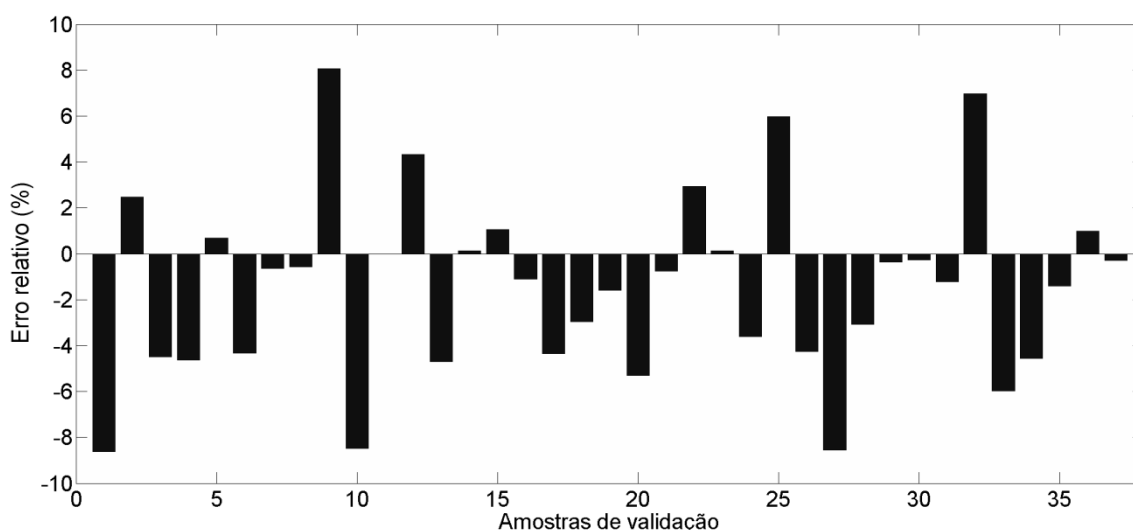


Figura 16 – Erros relativos das amostras de validação para o modelo de determinação de gordura

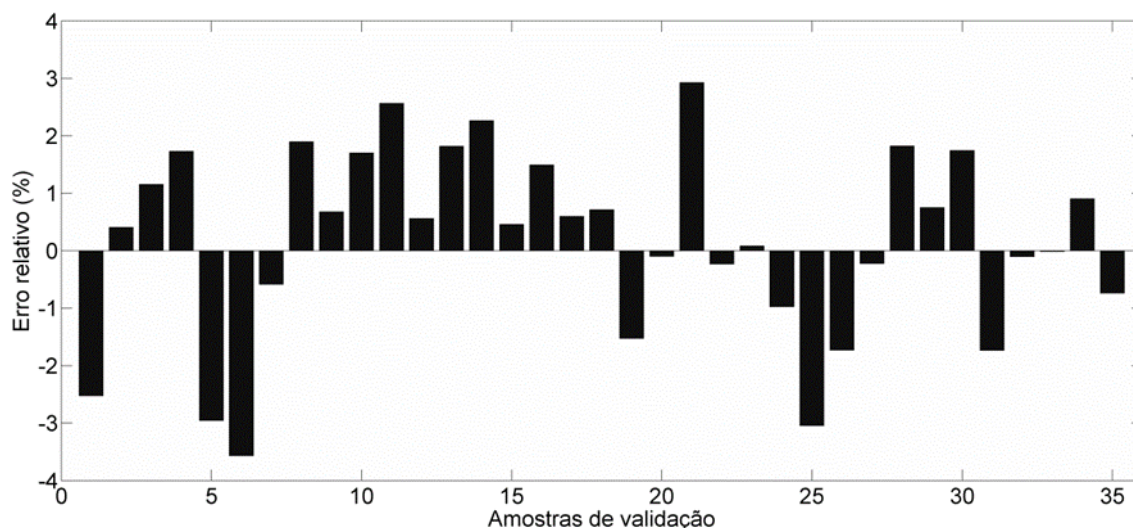


Figura 17 - Erros relativos das amostras de validação para o modelo de determinação de umidade

As precisões sob condições de repetitividade e precisão intermediária, expressas como valores de desvio padrão relativo (DPR_r e DPR_R , respectivamente), foram estimadas por análise de variância dos valores preditos para 18 replicatas de amostras em três níveis de concentração distintos.⁴⁷ Os valores de DPR_R variaram entre 3,1 e 8,5% para gordura e entre 1,6 e 2,2% para umidade, e os valores de DPR_r variaram entre 2,6 e 3,6% para gordura e entre 1,4 e 1,7% para umidade. Todos os valores estimados para o modelo da umidade estão abaixo do limite máximo, estimado através da equação de Horwitz, enquanto todos os valores estimados para o modelo da gordura estão acima do limite estabelecido.⁴⁰ Apesar de ser o método mais comum para se estimar o limite de precisão de métodos analíticos, e de ter sido utilizado neste trabalho, ele não é aplicável a todo tipo de análise. Segundo os autores, a equação de Horwitz não deve ser aplicada a determinações que são métodos-dependentes, como a determinação de gordura, umidade, ou cinzas, muito comuns em análises de alimentos. Por isso, mesmo o modelo de determinação de gordura não atendendo aos limites determinados pela equação, foi dada sequência à validação dos dois modelos.⁹⁷

Considerando-se os estudos de linearidade e precisão, os intervalos de trabalho dos métodos foram estabelecido como 38,7 - 58,0% p/p para a gordura e 41,5 - 55,1% p/p para a umidade. É importante ressaltar que estes intervalos

não são necessariamente idênticas às faixas úteis de calibração. Enquanto a calibração pode cobrir uma vasta gama de concentrações, o restante da validação abrange um leque mais restrito.⁴⁰

3.1.4.3.3. Seletividade e sensibilidade analítica

Como discutido anteriormente, os valores de SEL não têm interesse prático para o controle de qualidade de um ponto de vista multivariado. As estimativas de SEL apenas indicam que 7 e 11% do sinal analítico foram utilizados para prever a gordura e a umidade, respectivamente.

Uma vez que os valores de SEN não são apropriados para comparação com outros métodos, o γ dos métodos (Tabela 2) foi também calculado com base no ϵ estimado de 0,005. O inverso da γ indicou que os métodos foram capazes de discriminar diferenças de concentração mínimas de 0,6% para gordura e 0,4% para umidade, considerando o ruído aleatório instrumental como a única fonte de erros.

3.1.4.3.4. Robustez

Dois fatores de interesse prático foram escolhidos para a avaliação de robustez. O objetivo foi verificar se seria possível substituir as placas fornecidas pelo fabricante do equipamento por placas de petri convencionais, menos dispendiosas para as medidas de reflectância. E também se um pré-tratamento mais simples da amostra, sem trituração, proporcionaria os mesmos resultados. Um planejamento fatorial completo 2^2 (2 fatores em 2 níveis), em triplicata, indicou que os dois fatores estudados não apresentaram efeito significativo sobre os resultados no nível de confiança de 95%. Como pode ser visto na Figura 18, todos os efeitos estimados estão abaixo do valor crítico de 2,31, o que indica que o tipo de processamento e o tipo de placa utilizados não influenciam o resultado obtido. Portanto, para novos modelos, placas de petri comuns e amostras fatiadas poderão ser utilizadas.

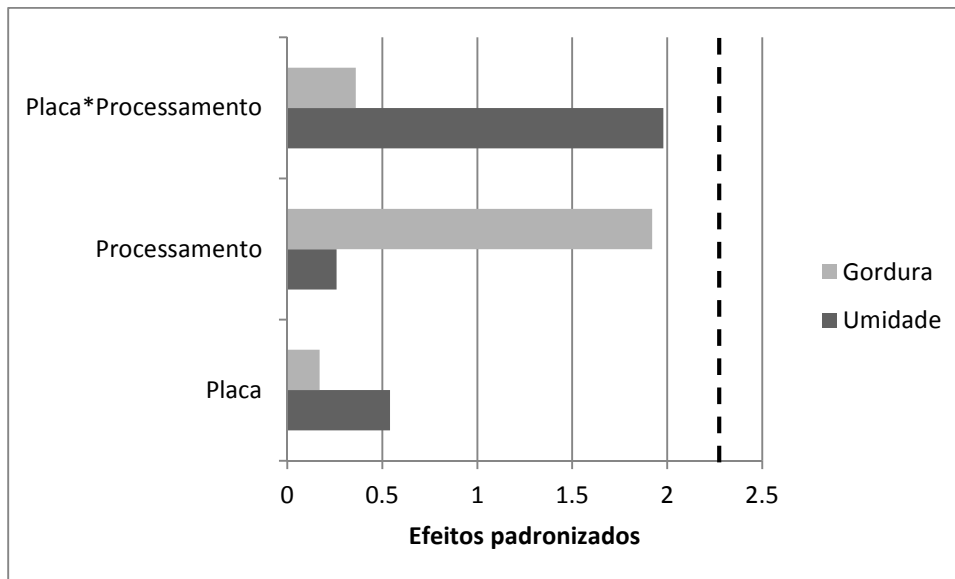


Figura 18 – Efeitos do teste de robustez, de acordo com um planejamento 2^2

3.1.4.3.5. Viés e RPD

O viés foi calculado apenas com base nas amostras de validação e as suas estimativas para ambos os modelos são mostradas na Tabela 2. Com os valores de viés e SDV, um valor de $t = 0,005$ foi estimado para o modelo de umidade, o que está abaixo do t crítico de 2,026 com 37 graus de liberdade e no nível de confiança de 95%. Para o modelo de gordura, $t = 2,426$ foi estimado, o que pode indicar a ausência de um viés estatisticamente significativo apenas a 98% de confiança ($t_{\text{crítico}}=2,431$).

De acordo com a literatura,⁵⁶ valores de RPD maiores que 2,4 são considerados desejáveis para boas equações de calibração e valores entre 2,4 e 1,5 são considerados satisfatórios, enquanto modelos com RPD inferior a 1,5 são considerados inadequados. Assim, os valores RPD para o modelo de umidade foram considerados bons, enquanto os valores de RPD para o modelo de gordura foram considerados satisfatórios (Tabela 2).

3.1.4.3.6 Erro amostral específico de previsão (EAEP)

Os EAEP variaram entre $\pm 1,1\%$ e $\pm 1,3\%$ para umidade e entre $\pm 2,5\%$ e $\pm 2,8\%$ para gordura. Comparando os valores preditos, juntamente com os intervalos de erro estimados, com os valores de referência, há uma concordância de 75% para o modelo de predição de gordura, enquanto a concordância para o modelo da umidade é de quase 90% (Figuras 19 e 20). Isso se deve

principalmente à exatidão dos métodos de referência. Como já foi discutido anteriormente, o método gravimétrico de dessecação em estufa é mais exato que o método de Gerber.

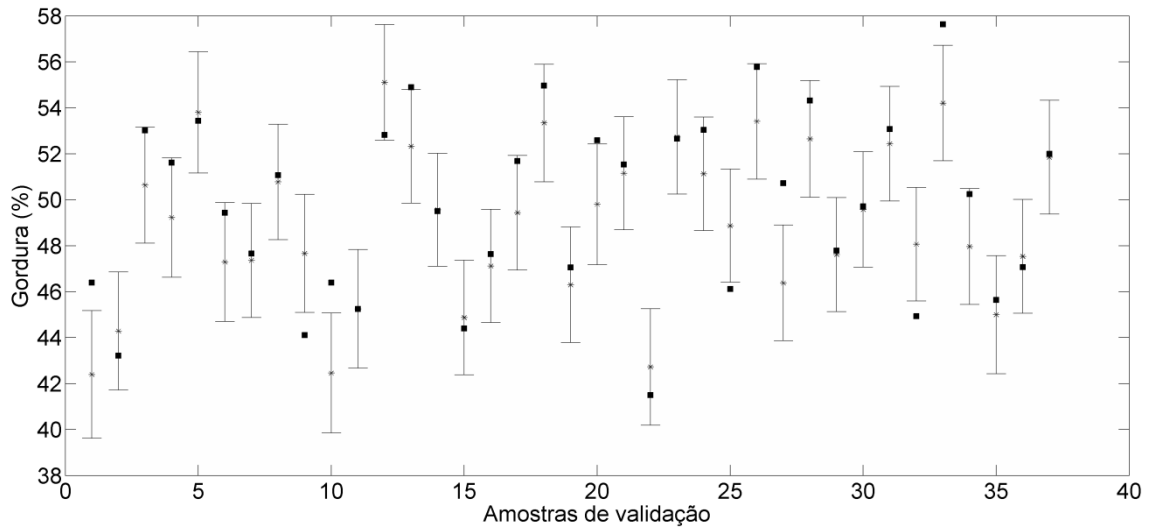


Figura 19 - Erro amostral específico de predição das amostras de validação para o modelo de determinação de gordura - valores de referência (●) e valores preditos (*)

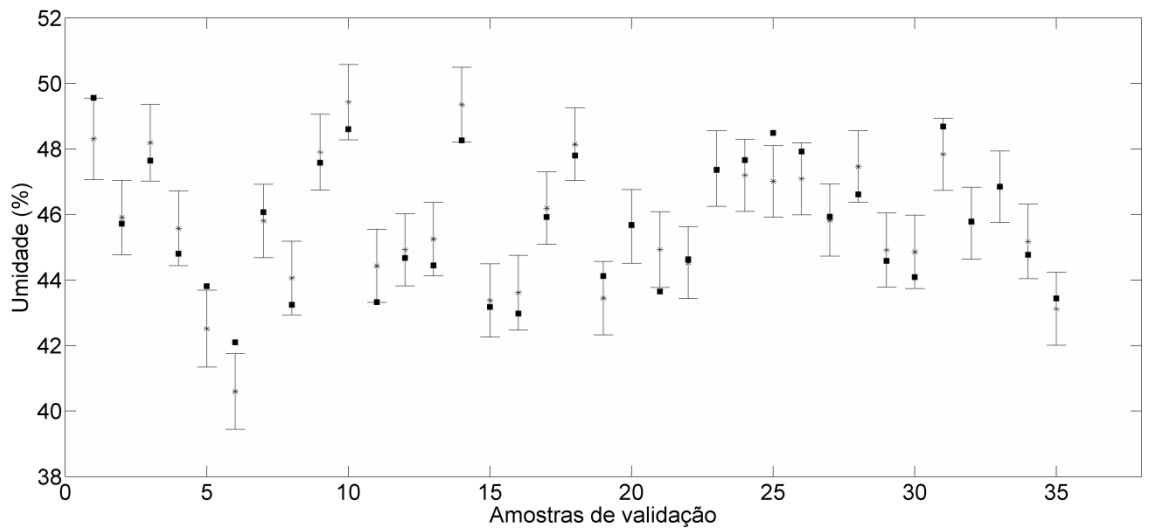


Figura 20 - Erro amostral específico de predição das amostras de validação para o modelo de determinação de umidade - valores de referência (●) e valores preditos (*)

3.1.4.3.7 Monitoramento pós modelo

Uma vez que os métodos foram validados, aprovados e implantados no IMA para o controle de qualidade de queijos muçarela, surgiu a necessidade de que fossem monitorados, a fim de avaliar as suas estabilidades. Durante um período de cerca de um ano, em 2011, 53 amostras foram coletadas e analisadas pelo método de referência e pelos modelos PLS desenvolvidos. Os resultados são mostrados como cartas de controle. O eixo y representa o erro relativo para cada amostra e o eixo x representa cada amostra em função da data de análise (Figuras 21 e 22). Os limites indicados nestas cartas de controle são baseados nas diretrizes brasileiras,³⁶ que estabelecem os limites de aceitação de erros para o monitoramento da estabilidade de métodos quantitativos.

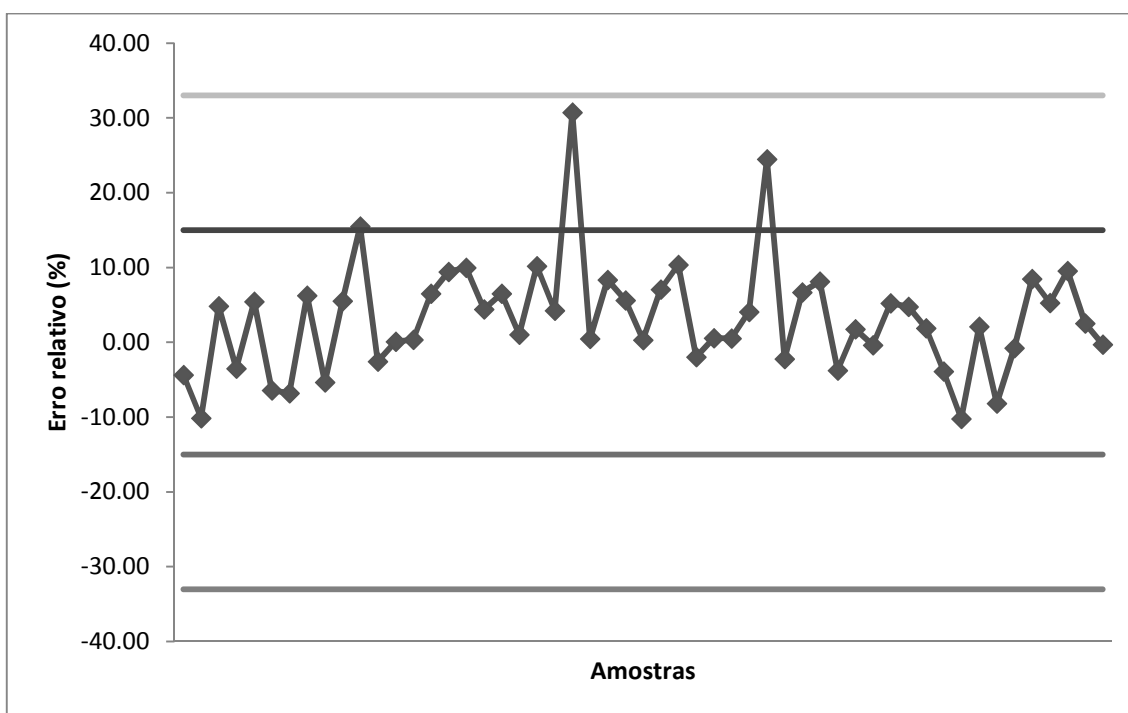


Figura 21 – Carta de controle para monitorar a estabilidade do modelo de determinação de gordura

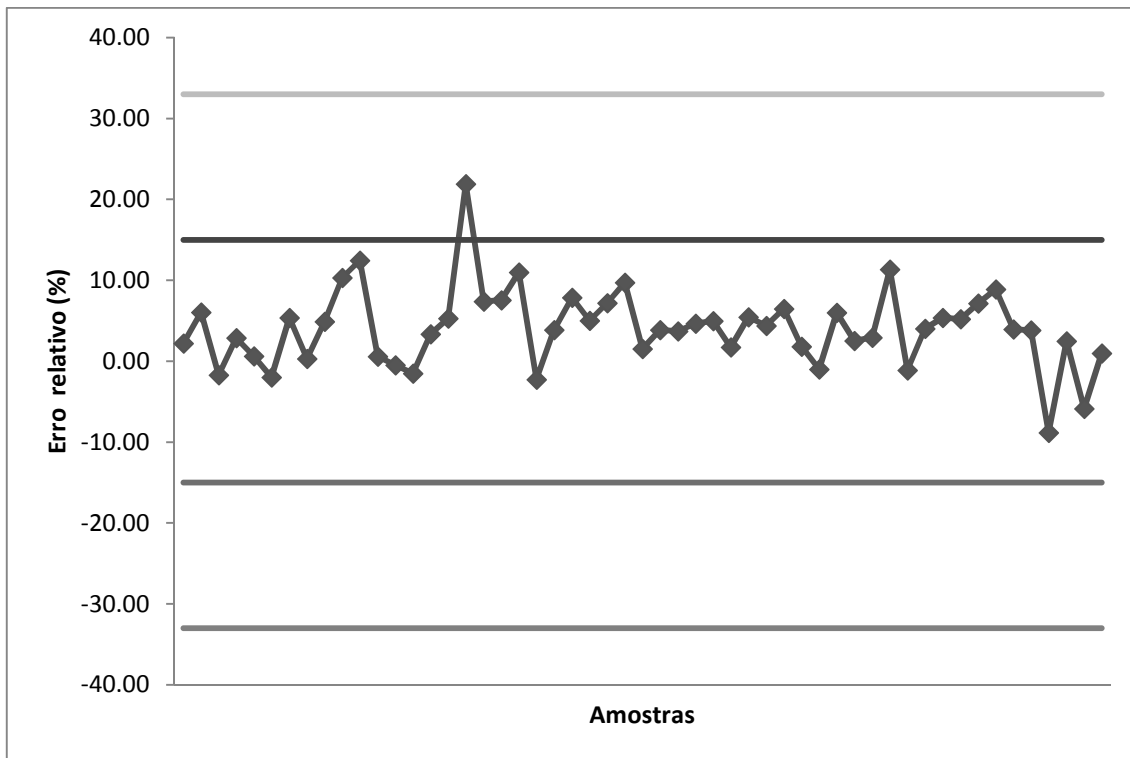


Figura 22 - Carta de controle para monitorar a estabilidade do modelo de determinação de umidade

De acordo com essa normatização, dois terços das amostras podem apresentar erros relativos dentro da faixa de $\pm 15\%$ (limites de alerta), enquanto um terço pode ter erros máximos admissíveis no limite máximo de $\pm 33\%$ (limites de ação). Uma vez que apenas três amostras (5,6% do total) para o modelo de gordura e uma amostra (1,9%) para o modelo de umidade apresentaram erros relativos entre 15% e 33%, ambos os modelos foram considerados satisfatoriamente estáveis.

3.2. Determinação de umidade em queijos minas artesanal utilizando NIR e PLS

A produção de queijos artesanais no Brasil iniciou-se no período colonial por portugueses que traziam consigo rebanhos bovinos. Os animais tinham parte de sua escassa produção leiteira destinada à fabricação de um queijo fresco, semelhante ao da Serra de Estrela, de Portugal. A diferença entre o queijo português e o produzido no Brasil referia-se ao tipo de material usado para coagulação do leite. Em Portugal, eram utilizados extratos de flores e brotos de

cardo silvestre (*Cynara cardunculus L.*), que no Brasil foram substituídos por estômago seco e salgado de mocó, bezerro ou cabrito.⁹⁸

No Brasil, são considerados como artesanais os queijos Minas, produzidos nas regiões de Araxá, Serra da Canastra, Cerrado, Serro e Campo das Vertentes, no estado de Minas Gerais; o queijo de Coalho e o queijo de Manteiga, produzidos na Região Nordeste; o queijo Serrano e o queijo Colonial, produzidos na Região Sul; e o queijo Caipira, produzido no estado do Mato Grosso do Sul. As estatísticas sobre a produção e consumo de queijos artesanais brasileiros são escassas.⁹⁸

Queijo minas é um queijo tradicional do brasileiro, sendo o queijo mais consumido em todo o país. Ele pode ser classificado em dois tipos diferentes: o Queijo minas artesanal, produzido apenas no estado de Minas Gerais, e o Queijo minas padrão, produzido em todo o país. As leis federais apenas reconhecem o queijo minas padrão, de modo que o queijo artesanal só pode ser comercializado dentro do Estado de Minas Gerais.⁹⁹

De acordo com as leis do Estado de Minas Gerais, para ser considerado queijo artesanal, alguns pré-requisitos devem ser cumpridos. O leite utilizado na produção de queijo deve ser da mesma fazenda onde o queijo será produzido, e este leite não pode ser pasteurizado. Em relação às propriedades físico-químicas, o teor de umidade não deve ser superior a 45,9%. O método oficial para determinação de umidade é a dessecação em estufa.^{100,101}

3.2.1 Objetivos

O objetivo deste trabalho foi desenvolver um método simples, robusto e rápido para determinação do teor de umidade em queijos minas artesanal usando NIRS e PLS. Esse método deve ser usado para o controle de qualidade destes queijos no IMA (Instituto Mineiro de Agropecuária). Este trabalho também tem como objetivos a realização das análises sob o ambiente do controle de qualidade, incluindo uma validação analítica e a avaliação permanente do desempenho do método através do monitoramento da estabilidade pós-modelo.

3.2.2 Material e métodos

3.2.2.1 Equipamentos e Softwares

Para a obtenção dos espectros e tratamento dos dados, foram utilizados, os mesmos equipamentos e *softwares* citados na sessão 3.3.1.

3.2.2.2 Materiais, reagentes e amostras

Noventa e cinco amostras de queijo minas foram coletadas entre 2009 e 2010. As amostras foram mantidas em temperatura inferior a 4 °C e analisadas após um período de 30 minutos, para que atingissem a temperatura ambiente. As condições para o preparo das amostras e a aquisição dos espectros foram similares às utilizadas para as amostras de muçarela, descritas na sessão 3.1.3.2.

O método de análise de referência foi realizado de acordo com a literatura. O teor de umidade total nos queijos foi determinado utilizando-se o método gravimétrico, sem repetições.⁷⁹ Três gramas de amostra foram pesadas e misturadas homogeneamente com a mesma quantidade de areia fina pré-seca, a fim de se evitar a formação de crosta e para acelerar o processo de secagem. As amostras foram, em seguida, secas a 105 °C, até atingirem peso constante.

3.2.3 Resultados e discussão

3.2.3.1 Desenvolvimento do modelo PLS

A Figura 23 mostra todos os espectros para as 95 amostras de queijo minas artesanal. É possível ver que as bandas mais intensas e largas estão por volta de 1450 e 1900 nm, o que corresponde ao primeiro sobretom de estiramento de ligações O-H e à banda de combinação de estiramento e deformação angular de O-H, respectivamente.^{83,84} As amostras foram divididas em um conjunto de calibração (63 amostras) e um conjunto de validação (32 amostras), utilizando o algoritmo de Kennard – Stone.⁹³

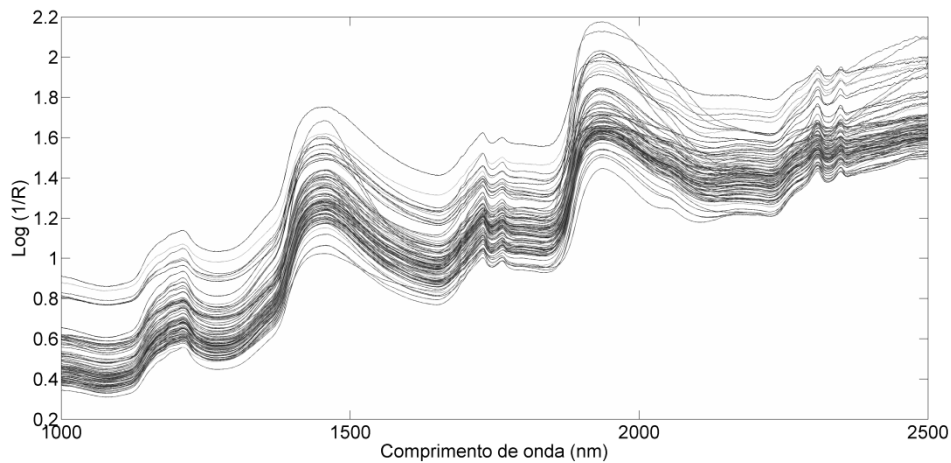


Figura 23 – Espectros das amostras de queijo minas artesanal

Após a separação dos grupos, os espectros foram pré-tratados para remover artefatos físicos. Como pode ser visto na Figura 23, desvios de linha base não lineares estão presentes nos espectros. Assim, para corrigi-los, a correção do espalhamento multiplicativo (MSC) foi usada, antes de centrar os dados na média. Os espectros pré-processados podem ser vistos na Figura 24. O melhor número de variáveis latentes foi escolhido baseado no menor valor de RMSECV, calculado com base na validação cruzada por blocos contínuos (7 divisões).

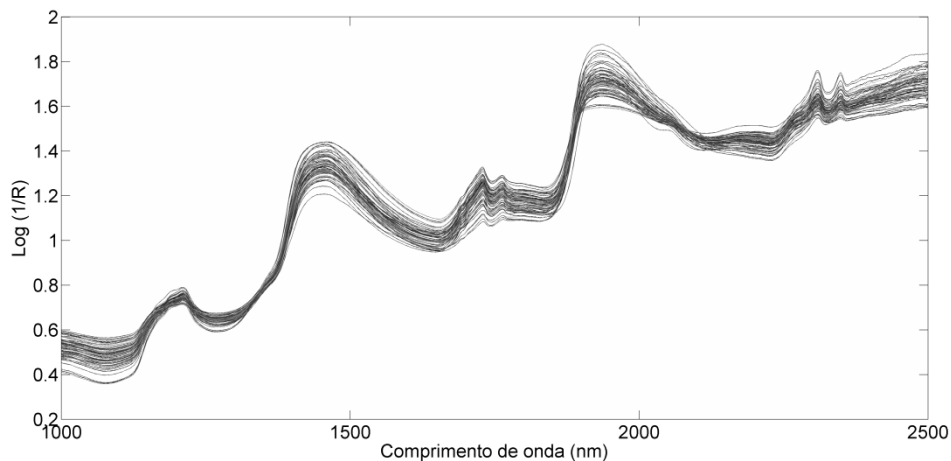


Figura 24 – Espectros de queijo minas artesanal pré-processados com MSC

Ao observar os coeficientes de regressão para o modelo (Figura 24), é possível observar, como esperado, que as regiões mais importantes estão relacionadas com vibrações de O-H, com os maiores valores positivos próximos de 1450 e 1900 nm. Também é interessante destacar os dois picos negativos, em 1200 e 1730 nm. Estes picos correspondem ao segundo e primeiro

sobretons, respectivamente, de ligações C-H,⁹⁰ as quais podem ser relacionadas aos teores de carboidratos, gorduras e proteínas no queijo, que estão inversamente relacionados com o teor de umidade.

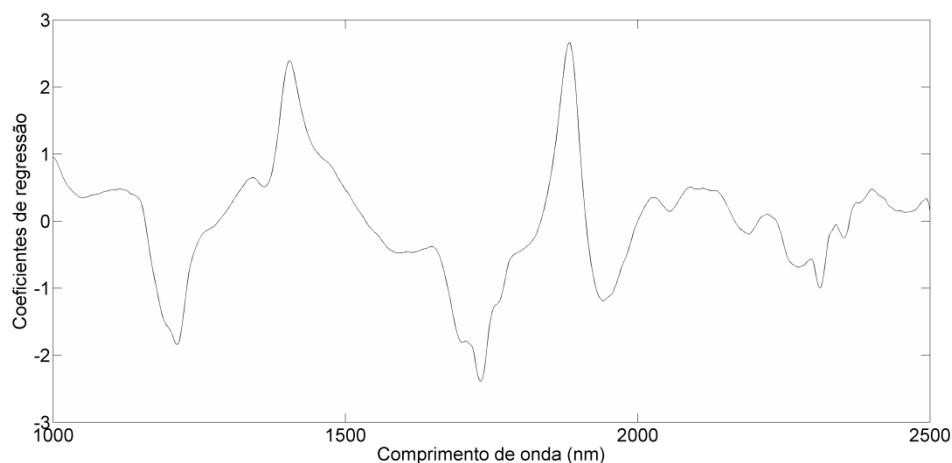


Figura 25 – Coeficientes de regressão para o modelo de determinação de umidade em queijo minas artesanal

3.2.3.2 Detecção de amostras anômalas e otimização do modelo

Após o desenvolvimento do modelo, o mesmo foi otimizado tal como descrito anteriormente, utilizando um nível de confiança de 99%. Os resultados estão resumidos na Tabela 3

Tabela 3 - Otimização dos modelos PLS para determinação de umidade em queijo minas artesanal pela detecção de *amostras anômalas* (modelo final está destacado em negrito)

Modelos	1	2	3
Amostras de calibração	63	58	55
Amostras de validação	32	32	27
Variáveis latentes	3	6	7
RMSEC (%)	3,1	2,0	1,5
RMSEP (%)	5,2	5,0	2,0

Oito *outliers* foram detectados no grupo de calibração, o que corresponde a 12% do total das amostras. Deste grupo, dois foram removidos por apresentarem alto *leverage*, um por alto resíduo espectral e cinco por altos resíduos em Y. No conjunto de validação, cinco *outliers* foram detectados (16% do total de amostras), todos eles por altos resíduos em Y. O modelo final otimizado foi construído com 55 amostras no conjunto de calibração e 27 no conjunto de validação, usando 7 VLs e explicando 99,5% da variância em X e 95,5% da variância em Y.

3.2.3.3 Validação analítica

A Tabela 4 resume os parâmetros estimados para cada FOM, os quais serão discutidos especificamente nas subseções seguintes.

Para confirmar esta aleatoriedade, alguns testes estatísticos foram utilizados. Os resíduos foram avaliados para distribuição normal (RJ), homocedasticidade (BF) e ausência de autocorrelação (DW), todos eles com um nível de confiança de 95%. O resultado de DW, apresentado na tabela 4, está dentro dos limites de aceitação (1,5-2,5), o que assegura que não há dependência entre os resíduos. Uma vez que o ajuste linear foi considerado um pressuposto válido, os parâmetros da curva de valores referência contra valores preditos foram calculados. A linearidade também pode ser expressa através da curva de calibração pseudounivariada (Figura 27).

3.2.3.3.1 Linearidade

A linearidade foi estimada a partir dos resíduos do ajuste dos valores de referência contra os valores preditos. Como pode ser observado, não há uma tendência sistemática visível na distribuição residual (Figura 26).

Tabela 4 - Figuras de mérito estimadas para o modelo PLS de determinação de umidade em queijo minas artesanal

Figuras de Mérito	Paramêtros	Valores
Veracidade (n=27)	Erro relativo médio	-0,7%
Precisão (n=18)	DPR repetitividade ^a (DPR _r)	1,4 % (1,6 %)
		1,4 % (1,5 %)
	DPR precisão intermediária ^b (DPR _R)	1,5 % (1,6 %)
		1,6 % (2,3 %)
		1,4 % (2,3 %)
	1,7 % (2,4 %)	
Linearidade	Durbin-Watson	2,0
	Inclinação	0,95 ± 0,06
	Intercepto	1,86 ± 2,40
	Coeficiente de correlação (r) ^c	0,9773
Faixa de trabalho		28,5 a 55,5 %
Seletividade do modelo		0,120
Sensibilidade do modelo		0,058 ^d
Sensibilidade analítica (γ)		10,34 % ⁻¹
γ ⁻¹		0,1%
<i>Bias</i>		0,820±1,902 %
RPD	RPD calibração	3,4
	RPD validação	2,1

^a Resultados para três amostras em três diferentes concentrações analisadas no mesmo dia por um mesmo analista, em sextuplicata. Valores limites de DPR_r (entre parênteses) calculados através da equação de Horwitz para cada nível de concentração estudado.

^b Resultados para três amostras em três diferentes concentrações em dias diferentes por analistas diferentes, em sextuplicata. Valores limites de DPR_R (entre parênteses) calculados através da equação de Horwitz para cada nível de concentração estudado.

^c Parâmetros do ajuste para a curva de valores de referência vs. valores preditos para as amostras de calibração

^d Valores expressos como a razão entre log (1/R) e %

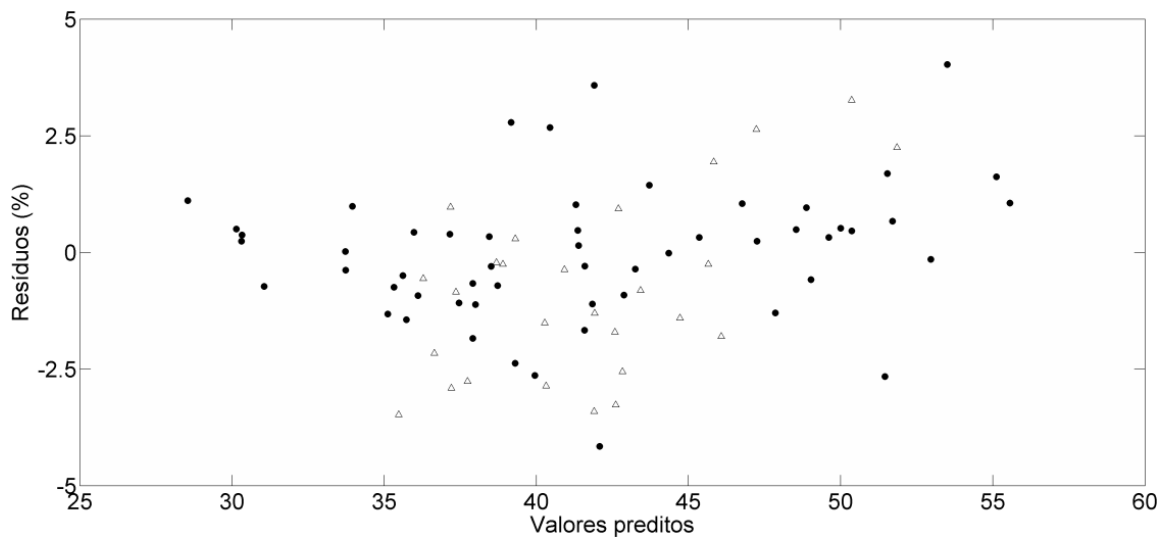


Figura 26 - Gráfico de resíduos vs. valores preditos para o modelo de determinação de umidade – amostras de calibração (●) e de validação (Δ)

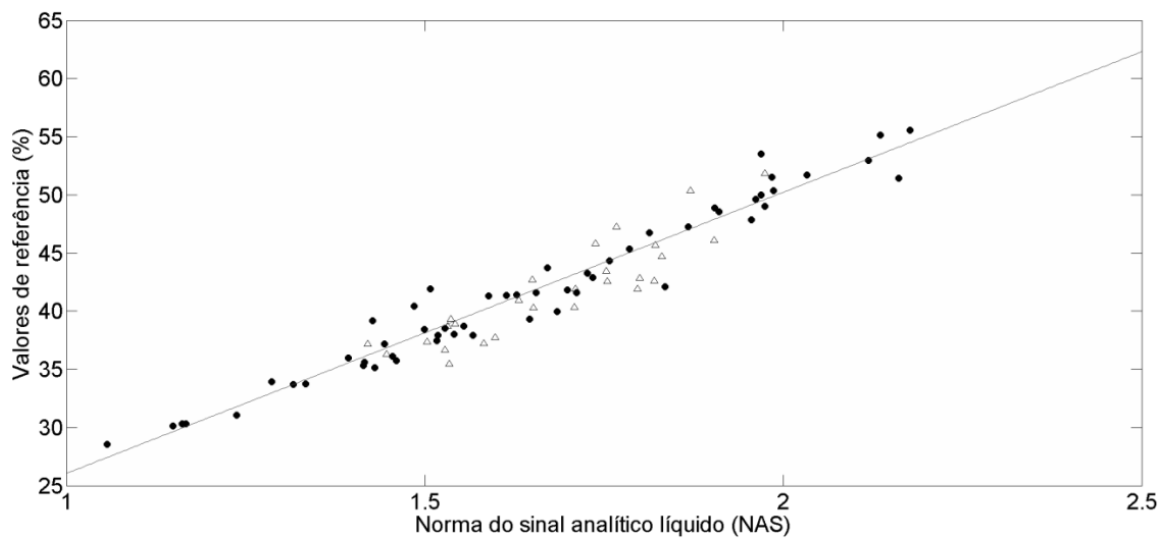


Figura 27 - Curva de valores de referência vs. norma do NAS para o modelo de determinação de umidade - amostras de calibração (●) e de validação (Δ)

3.2.3.3.2 Exatidão e faixa de trabalho

Para este modelo, o RMSEP foi de 2,0% (Tabela 4), e os erros relativos variaram entre -9,8 e 6,5%, com valor médio de -0,7% (Figura 27). De acordo com a AOAC⁴¹, os limites de aceitação de erros relativos são -3%/3%, de modo que pode ser considerado que o método fornece resultados verdadeiros. Os valores de precisão sob condições de repetitividade e precisão intermediária, expressos como desvio padrão relativo (DPR_r e DPR_R, respectivamente), foram

estimadas por análise de variância dos valores preditos para 18 replicatas de amostras em três níveis de concentração distintos.⁴⁷ Os valores de DPR_R variaram entre 1,4 e 1,7%, e os valores de DPR_r variaram entre 1,4 e 1,5%. Todos os valores estimados para o modelo da umidade estão abaixo do limite máximo, calculado através da equação de Horwitz.⁴⁰ Os resultados de veracidade e precisão permitem assegurar que o método é exato.

Considerando-se os estudos de linearidade e precisão, a faixa de trabalho do método foi estabelecido como 28,5 - 55,5% p/p.

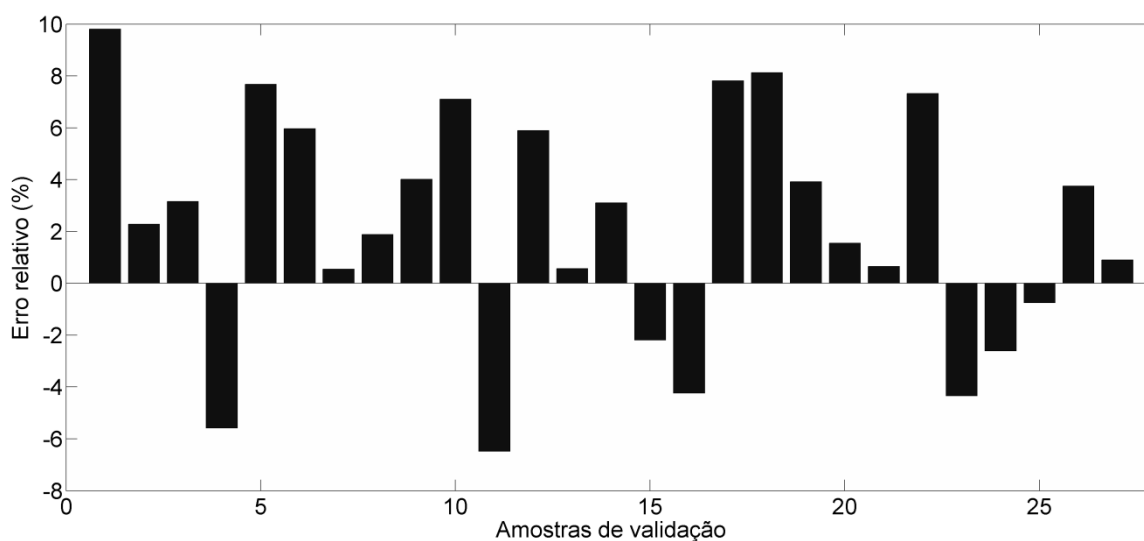


Figura 28 - Erros relativos das amostras de validação para o modelo de determinação de umidade

3.2.3.3.3 Seletividade e sensibilidade analítica

Como discutido antes, a SEL não tem interesse prático em modelos quimiométricos. A sua estimativa apenas nos dá uma ideia de quanta informação espectral foi utilizada para a construção do modelo. Para este modelo, 12% do sinal analítico foi utilizado na predição da umidade.

A sensibilidade analítica (γ), calculada usando a SEN e o ruído instrumental (ϵ), é uma FOM útil, que permite a comparação entre a sensibilidade de diferentes técnicas, e o inverso da γ (γ^{-1}) indica a concentração mínima que o modelo pode discriminar, considerando o ruído instrumental como a única fonte de erros. O γ^{-1} calculado para este modelo foi de 0,1% (Tabela 4).

3.2.3.3.4 Viés e RPD

O viés foi calculado com base apenas em amostras de validação, utilizando 27 graus de liberdade, e é apresentada na tabela 4. O resultado de $0,820 \pm 1,920$ indica ausência de erros sistemáticos, no nível de confiança de 95%. Levando em conta os limites estabelecidos para RPD na literatura,^{56,57} o presente modelo pode ser considerado adequado para quantificação (Tabela 4).

3.2.3.3.5. Erro amostral específico de predição (EAEP)

Neste cálculo, a incerteza do método de referência, 0,1%, foi obtida de dados históricos. De acordo com a Equação 29, 9 pseudo graus de liberdade foram usados no cálculo dos intervalos de confiança ($t = 2,03$). O EAEP estimado variou entre 2,5 e 2,7%. A Figura 29 mostra os valores preditos com o intervalo de EAPP e os valores de referência para cada amostra de validação. De um total de 27 amostras, sete estavam fora dos limites do intervalo. Este elevado número de amostras fora dos limites pode ser devido à capacidade do método de referência em determinar apenas água livre na amostra, e o NIRS detecta toda a água, até mesmo a quimicamente ligada. Talvez, devido a este fato, seis das sete amostras citadas acima têm valores de referência menores do que os valores preditos.

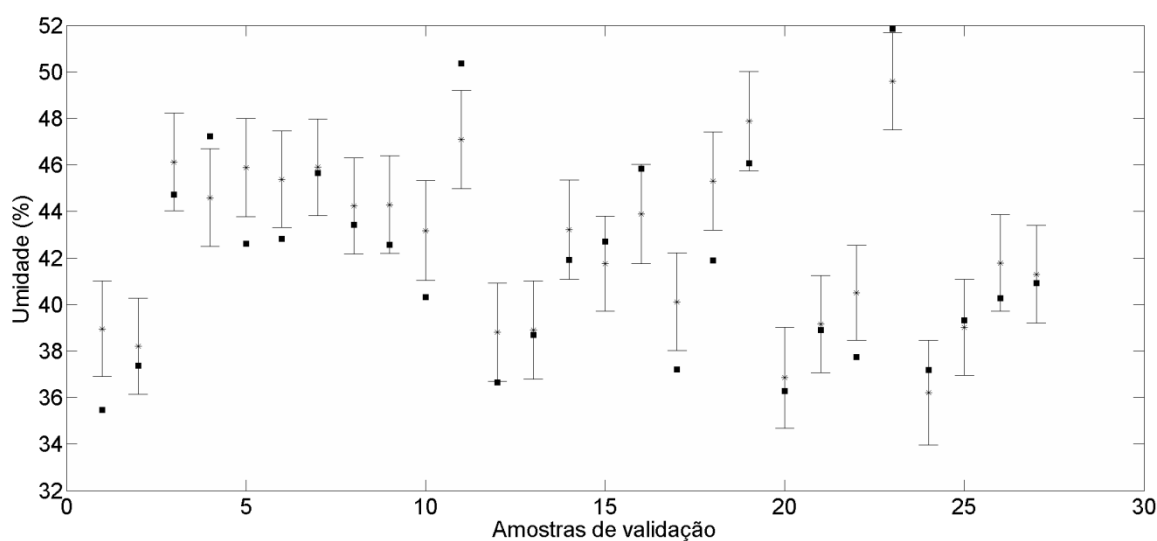


Figura 29 – Erro amostral específico de predição para as amostras de validação – valores de referência (●) e valores preditos (*)

3.2.3.3.6 Monitoramento pós-modelo

Depois que o método foi validado e adotado como análise de rotina para determinação de umidade em queijo minas artesanal, o mesmo foi monitorado a fim de avaliar a sua estabilidade. Durante um ano, pelo menos uma amostra por semana foi analisada em ambos os métodos, de referência e NIRS, e a diferença foi avaliada através dos erros relativos. Os resultados são mostrados na Figura 30, através de uma carta de controle. O limite de alerta ($\pm 15\%$) e o limite de ação ($\pm 33\%$) foram estabelecidos de acordo com as diretrizes brasileiras [39]. Para ser considerado estável, o modelo pode apresentar dois terços das amostras analisadas com erros relativos dentro do limite de atenção ($\pm 15\%$), enquanto apenas um terço pode estar entre o limite de atenção e o limite de ação ($\pm 33\%$).

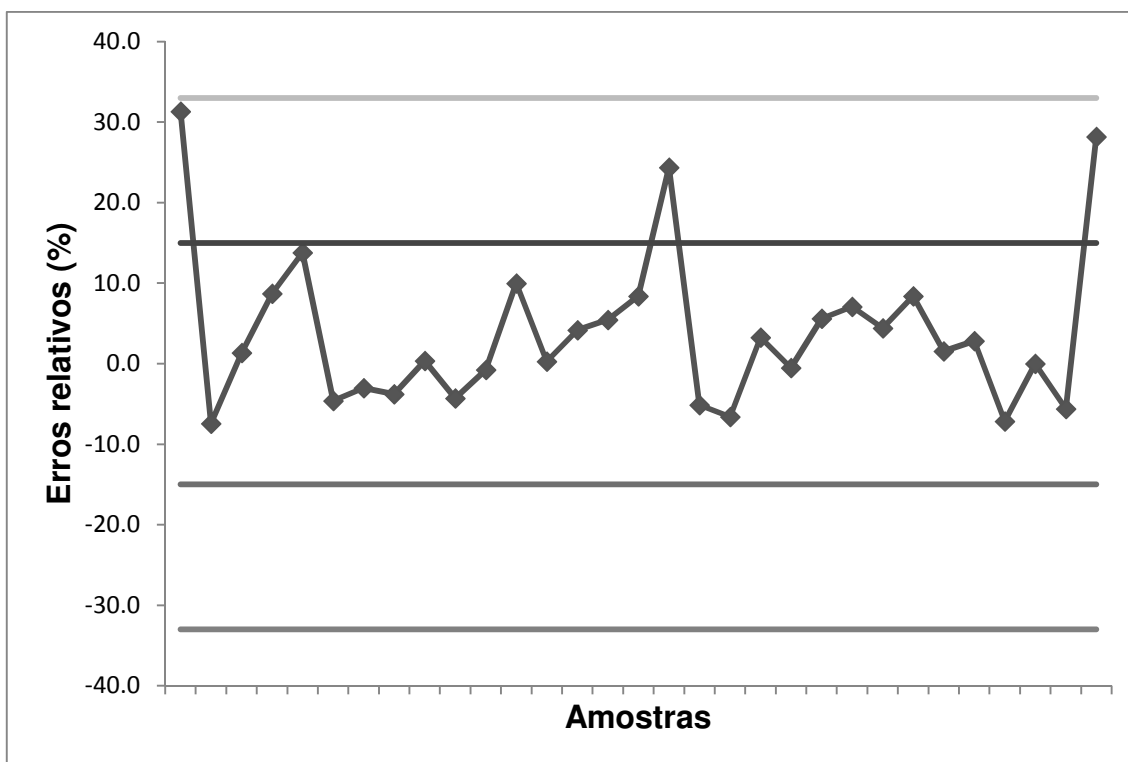


Figura 30 – Carta de controle da estabilidade do modelo para determinação de umidade em queijo minas artesanal

Apenas três amostras ficaram acima do limite de atenção (10% do total de amostras analisadas), o que indica que o modelo permaneceu estável durante um ano após a sua validação.

3.3 Conclusão

A implementação de modelos multivariados para o controle de qualidade de queijos minas e muçarela se mostrou uma alternativa viável aos métodos tradicionais. Ambas as aplicações foram desenvolvidas em condições reais de análise, sem alterar os protocolos e a rotina convencional do laboratório.

Estes métodos apresentaram vantagens sobre os métodos tradicionais usados na rotina, tais como baixo custo, processo simplificado, sem necessidade de reagentes e sem gerar resíduos químicos. A principal vantagem dos métodos desenvolvidos foi a rapidez das análises, uma vez que apenas cerca de três minutos foram gastos para a determinação de ambos os parâmetros. Assim, os métodos NIRS foram cerca de 30 vezes mais rápidos do que o método de Gerber para a determinação de gordura, e 80 vezes mais rápidos do que o método gravimétrico para a determinação de umidade.

Todos os métodos foram validados em conformidade com as diretrizes brasileiras e internacionais, sendo considerados lineares, precisos, sensíveis e robustos. Além disso, após os métodos serem adotados, seus desempenhos foram monitorados durante cerca de um ano empregando de cartas de controle, demonstrando que eles continuaram a fornecer previsões aceitáveis.

Além da redução de custos com a utilização da areia purificada pré-seca, que é um produto com preço elevado (R\$ 1000 por quilo, como são consumidos 3 g por medida, estima-se que só com a economia deste reagente deixe-se de gastar cerca de R\$ 3,10 em cada análise), e da eliminação da necessidade de se trabalhar com ácido sulfúrico concentrado, um reagente potencialmente perigoso, a substituição dos métodos tradicionais de análise permitiu uma redução do tempo total de análise e eliminou a geração de resíduos.

Finalmente, mais do que estes métodos específicos, esta aplicação teve como objetivo apresentar uma completa e robusta metodologia multivariada. Esta metodologia empregou o conceito de sinal do analito líquido e harmonizou a validação tradicional univariada com os requisitos dos métodos multivariados

de calibração, o que é essencial para a implementação de novos métodos NIRS na rotina de análise para controle de qualidade dos alimentos.

4. Desenvolvimento de um método de triagem para a detecção simultânea de cinco adulterantes em leite cru utilizando infravermelho médio e PLS-DA

4.1 Introdução

O leite pode ser definido como o fluido secretado pelas fêmeas de todas as espécies de mamíferos, com função primária de suprir a demanda nutricional do neonato, principalmente energia, aminoácidos essenciais e grupos amino para a biossíntese de aminoácidos não essenciais, ácidos graxos essenciais, vitaminas, elementos inorgânicos e água. Além disso, o leite também se tornou o componente majoritário em dietas humanas em diversas partes do planeta.¹⁰²

Em 2011, foram produzidos mais de 735 milhões de toneladas de leite no mundo, incluindo leites de vaca, búfala, cabra, ovelha e camela. Os EUA foi o país com a maior produção mundial (87,5 milhões de toneladas), enquanto o Brasil ocupa a quinta posição, com 31,7 milhões de toneladas. Entretanto, assim como a demanda e a produção de leite têm crescido, um número crescente de adulterações sofisticadas tem sido reportado em diferentes países.¹⁰³

A adição de adulterantes em leite com a intenção de aumentar os lucros não é um assunto novo, mas recentemente tem atraído mais atenção por todo o mundo. No Brasil, dois escândalos recentes de adulteração de leite foram descobertos em operações da Polícia Federal. Em 2007, a “Operação Ouro Branco” denunciou o uso de citrato de sódio, hidróxido de sódio, cloreto de sódio, sacarose, fosfatos, carbonatos, bicarbonatos e peróxido de hidrogênio para a correção de defeitos no leite, como acidez elevada e crescimento microbiano, além de mascarar a adição de água.¹⁰⁴ Em 2013, a “Operação Leite Compen\$ado” revelou a utilização de fertilizantes contendo ureia e formol para mascarar a adição de água em leite.¹⁰⁵ Souza *et al.*¹⁰⁶ analisaram 100 amostras de leite UHT comercializadas no Brasil e detectaram uma elevada taxa de não conformidades, com 55 % das amostras apresentando urina, 44 % com presença de formol, 30 % com peróxido de hidrogênio e 12 % com presença de cloretos. Atualmente, a detecção de adulterantes em leite é realizada através de métodos físico-químicos baseados em reações específicas e variações no ponto de

congelamento ou gravidade específica.^{107,108} Estes métodos são demorados, trabalhosos e podem não fornecer resultados precisos.

Uma vez que a adição de qualquer tipo de substância estranha ao leite cru é proibida no Brasil, em qualquer nível, somente métodos qualitativos são descritos nos compêndios oficiais para a avaliação de adulteração em leite. Por isso, é mais importante o desenvolvimento de métodos analíticos que detectem com confiabilidade os possíveis adulterantes.^{107,108}

Nos últimos anos, a utilização de metodologias quimiométricas de classificação supervisionada e/ou calibração multivariada combinadas com espectroscopia no infravermelho (médio ou próximo) tem permitido o desenvolvimento de métodos rápidos e não destrutivos para detectar e/ou quantificar adulterantes em leite, tais como ureia, urina sintética, leite sintético, soro de leite, peróxido de hidrogênio, água, melamina e nitrato de amônio.^{109–113} Um método alternativo, baseado em análise de imagens também foi aplicado para detectar e quantificar a adição de água e hidróxido de sódio em leite.¹²

Considerando as adulterações reportadas recentemente, a presença simultânea de diversos adulterantes é comum, mas a maioria dos métodos quimiométricos encontrados na literatura determinam apenas um ou no máximo dois adulterantes simultaneamente.^{109,112,113} Para esta aplicação, foram escolhidos cinco dos adulterantes mais comuns recentemente identificados pela Polícia Federal em amostras de leite brasileiro: água, amido, citrato de sódio, formol e sacarose. A água é o adulterante mais comum, utilizado para aumentar o volume total do leite por diluição, resultando conseqüentemente em uma diminuição da concentração dos componentes do leite, como proteínas, gordura e lactose. O amido é utilizado como espessante, aumentando também o teor de sólidos totais do leite adulterado. O citrato de sódio é utilizado como estabilizante e conservante, evitando a precipitação de nutrientes. O formol é utilizado para evitar o desenvolvimento e a contaminação microbiana do leite adulterado, uma vez que más condições de armazenamento e produção, além de excessiva manipulação do produto durante uma possível adulteração, podem elevar o risco de contaminação microbiana. A sacarose é utilizada para restaurar os valores normais de amostras de leite adulteradas nos testes de controle físico-químicos, além de melhorar as suas propriedades sensoriais.

4.2 Objetivos

Desenvolver um método de triagem rápido e não destrutivo para a detecção simultânea de até cinco adulterantes (água, amido citrato de sódio, formaldeído e sacarose) em leite bovino cru, utilizando espectroscopia no infravermelho médio com acessório de reflexão total atenuada (ATR) e PLS-DA.

Estimar algumas FOM específicas utilizadas em validação analítica qualitativa, tais como sensibilidade, especificidade, taxa de confiabilidade, acordância e concordância. Considerando que são poucos os trabalhos na literatura que implementaram uma validação analítica de métodos de classificação multivariados, este capítulo se propõe também a harmonizar os aspectos univariados presentes na maioria das normatizações com a realidade multivariada.

4.3 Materiais e métodos

4.3.1 Equipamentos e softwares

Os espectros de infravermelho foram adquiridos utilizando um espectrofotômetro Shimadzu IR Affinity1, usando um acessório de ATR equipado com uma cela de ZnSe. Os *softwares* utilizados foram os mesmos citados na sessão 3.3.1.

4.3.2 Amostras e reagentes

As amostras brancas de leite bovino cru foram fornecidas pela Fazenda Modelo da Escola de Veterinária da Universidade Federal de Minas Gerais (UFMG), localizada em Pedro Leopoldo/MG. Citrato de sódio, formol, amido, sacarose (todos de grau analítico) e água destilada foram utilizados como adulterantes.

4.3.3 Procedimento

O leite foi armazenado em garrafas de vidro âmbar a -18 °C, aquecidas em banho de água até temperatura ambiente para o preparo das amostras. Neste preparo, todos os adulterantes foram pesados em balança analítica ($\pm 0,0001$ g). Para as amostras contendo apenas um adulterante, água, sacarose, formol, amido ou citrato foram adicionados em concentrações variando de 0,5 a

10,0% p/v, com incrementos de 0,5%, resultando em 20 amostras adulteradas para cada analito. Para o preparo das amostras com dois ou mais adulterantes, um planejamento de misturas com cinco componentes foi empregado. Neste planejamento, as concentrações individuais dos componentes variaram entre 0,5 e 10,0 %, resultando em misturas de concentrações finais, somando todos os adulterantes presentes, iguais a 10,0 %. O planejamento proposto resultou em 71 amostras (30 amostras binárias, 30 amostras ternárias, 5 amostras quaternárias e 6 amostras pentenárias), cuja composição está discriminada na Tabela 5.

Todas as amostras foram analisadas diretamente, sem nenhum pré-tratamento, colocando-se 30 μL no acessório de ATR. Os espectros foram obtidos como médias de 20 varreduras, de 600 a 4000 cm^{-1} . Antes de cada medida, uma correção de linha de base foi realizada para evitar a interferência atmosférica e reduzir o ruído instrumental. Após a aquisição de cada espectro, o acessório foi limpo com acetona P.A.

Tabela 5– Planejamento de mistura com cinco componentes para a adulteração de leite com água, amido, citrato de sódio, formol e sacarose

Adulteração (% p/v)					
Amostras	Água	Amido	Citrato	Formol	Sacarose
1	0,0	5,0	2,5	0,0	2,5
2	2,5	0,0	0,0	5,0	2,5
3	2,5	0,0	5,0	0,0	2,5
4	7,5	2,5	0,0	0,0	0,0
5	5,0	0,0	0,0	2,5	2,5
6	2,5	0,0	2,5	2,5	2,5
7	0,0	0,0	7,5	2,5	0,0
8	0,0	0,0	2,5	0,0	7,5
9	6,0	1,0	1,0	1,0	1,0
10	0,0	2,5	5,0	2,5	0,0
11	5,0	0,0	0,0	5,0	0,0
12	1,0	1,0	1,0	6,0	1,0
13	0,0	5,0	5,0	0,0	0,0
14	0,0	2,5	2,5	0,0	5,0
15	0,0	2,5	0,0	7,5	0,0
16	0,0	0,0	0,0	5,0	5,0
17	2,5	5,0	0,0	0,0	2,5
18	0,0	2,5	0,0	0,0	7,5
19	2,5	0,0	0,0	0,0	7,5

20	2,5	2,5	0,0	0,0	5,0
21	0,0	0,0	5,0	5,0	0,0
22	0,0	2,5	7,5	0,0	0,0
23	0,0	5,0	0,0	5,0	0,0
24	2,5	0,0	7,5	0,0	0,0
25	2,5	2,5	2,5	0,0	2,5
26	2,5	5,0	0,0	2,5	0,0
27	5,0	0,0	0,0	0,0	5,0
28	5,0	2,5	2,5	0,0	0,0
29	0,0	5,0	0,0	0,0	5,0
30	0,0	2,5	2,5	5,0	0,0
31	0,0	0,0	0,0	2,5	7,5
32	1,0	1,0	6,0	1,0	1,0
33	2,5	7,5	0,0	0,0	0,0
34	2,5	0,0	0,0	2,5	5,0
35	0,0	7,5	0,0	0,0	2,5
36	7,5	0,0	0,0	0,0	2,5
37	0,0	0,0	2,5	7,5	0,0
38	5,0	2,5	0,0	2,5	0,0
39	0,0	2,0	2,0	2,0	2,0
40	0,0	0,0	2,5	5,0	2,5
41	0,0	0,0	0,0	7,5	2,5
42	1,0	1,0	1,0	1,0	6,0
43	2,5	2,5	0,0	5,0	0,0
44	0,0	0,0	2,5	2,5	5,0
45	2,5	5,0	2,5	0,0	0,0
46	0,0	0,0	5,0	0,0	5,0
47	5,0	0,0	5,0	0,0	0,0
48	0,0	0,0	5,0	2,5	2,5
49	0,0	5,0	2,5	2,5	0,0
50	0,0	2,5	2,5	2,5	2,5
51	7,5	0,0	0,0	2,5	0,0
52	0,0	7,5	0,0	2,5	0,0
53	2,5	0,0	5,0	2,5	0,0
54	5,0	0,0	2,5	0,0	2,5
55	2,5	2,5	2,5	2,5	0,0
56	5,0	2,5	0,0	0,0	2,5
57	0,0	7,5	2,5	0,0	0,0
58	0,0	2,5	0,0	5,0	2,5
59	5,0	5,0	0,0	0,0	0,0
60	0,0	0,0	7,5	0,0	2,5
61	7,5	0,0	2,5	0,0	0,0
62	2,5	2,5	0,0	2,5	2,5
63	1,0	6,0	1,0	1,0	1,0
64	0,0	2,5	0,0	2,5	5,0
65	0,0	2,5	5,0	0,0	2,5

66	5,0	0,0	2,5	2,5	0,0
67	2,5	2,5	5,0	0,0	0,0
68	2,5	0,0	2,5	5,0	0,0
69	2,5	0,0	0,0	7,5	0,0
70	2,5	0,0	2,5	0,0	5,0
71	0,0	5,0	0,0	2,5	2,5

4.4 Resultados e discussão

4.4.1 Espectros de infravermelho

Os espectros de todas as amostras estão mostrados na Figura 31. Pela observação destes espectros, é possível atribuir a banda mais larga e intensa, entre aproximadamente 3600 e 3000 cm^{-1} , a vibrações de estiramento de ligações O-H. Os dois picos pequenos próximos a 2800 cm^{-1} estão relacionados com o estiramento dos grupos CH_2 de ácidos graxos do leite. Outro pico, próximo de 1700 cm^{-1} , é característico de amida I e II, e pode ser relacionado ao conteúdo proteico do leite.^{110,111} Inicialmente, a região entre 2430 e 2230 cm^{-1} foi removida devido à interferência do CO_2 atmosférico. Mesmo com a correção atmosférica instrumental, algumas amostras apresentaram um pico característico nesta região espectral, a qual foi então removida de todas as amostras.

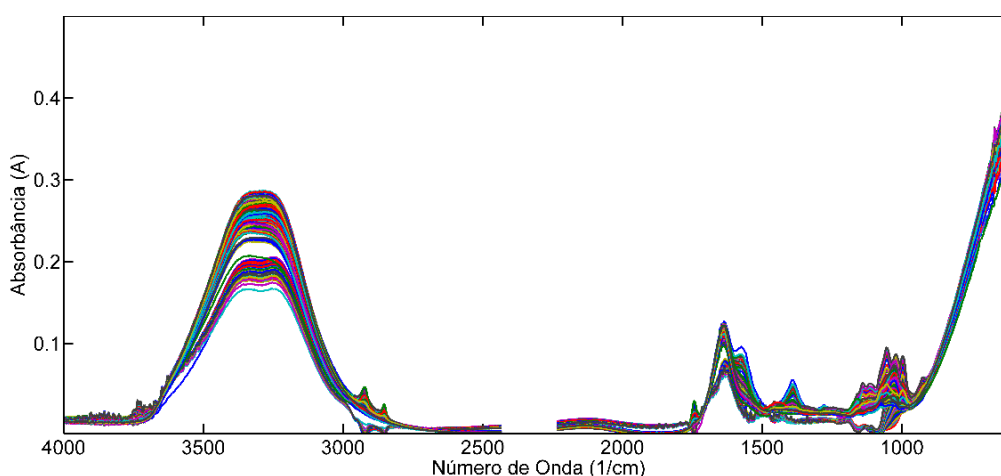
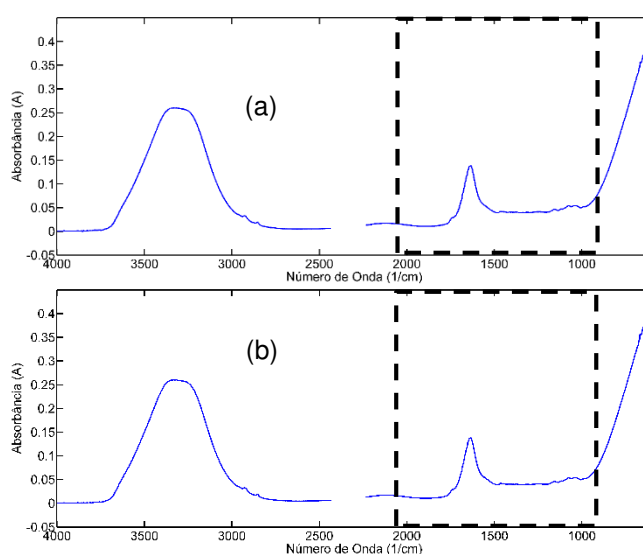


Figura 31 – Espectros de infravermelhos de todas as amostras de leite cru bovino (amostras brancas, adulteradas com um único adulterantes e adulteradas com as misturas de adulterantes)

A Figura 32(a) mostra um espectro de uma amostra de leite branca, sem a adição de nenhum adulterante. Como mencionado anteriormente, dois picos podem ser

notados, relacionados com os teores de água e proteína do leite. A Figura 32(b) mostra o espectro de uma amostra de leite adulterada com 10,0 % p/v de água, e nenhuma mudança visual significativa pôde ser notada no espectro. A Figura 32(c) mostra o espectro de uma amostra adulterada com 10,0 % p/v de amido. Em comparação com a Figura 32(a), é possível notar pequenas variações na região de impressão digital (*fingerprint*), entre 1200 e 900 cm^{-1} . Essas absorções podem ser associadas com vibrações da ligação glicosídica α -1,4, deformações angulares de C-H e estiramentos de C-O e C-C.¹¹⁴ A Figura 32(d) mostra o espectro de uma amostra adulterada com 10,0 % p/v de citrato de sódio. As duas bandas observadas próximas a 1400 e 1600 cm^{-1} podem ser relacionadas com os estiramentos simétrico e assimétrico de ligações C-O de grupos carboxilato. A Figura 32(e) mostra o espectro de uma amostra adulterada com 10,0 % p/v de formol. Um pico característico próximo a 1000 cm^{-1} pode ser destacado. Este pico está relacionado com as vibrações de estiramento e balanço (*wagging*) de grupos CH_2 do formol.⁵ A Figura 32 (f) mostra um espectro de uma amostra adulterada com 10,0 % p/v de sacarose. A presença de sacarose pode ser associada com o aparecimento de diversos picos na região de *fingerprint*, atribuídos aos estiramentos de C-H e C-O, estiramento e deformação angular de C=O de cetonas e estiramento e deformação de C-O.¹¹⁵



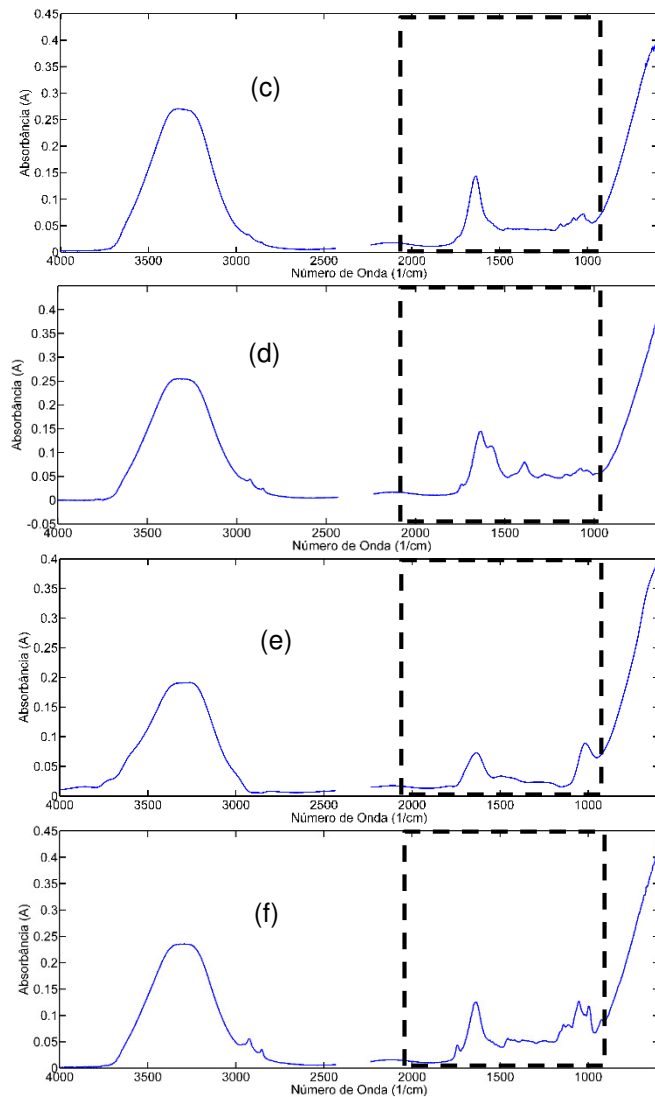


Figura 32 – Espectro de infravermelho médio de (a) amostra branca; (b) adulteração com água (10,0 % p/v); (c) adulteração com amido (10,0 % p/v); (d) adulteração com citrato de sódio (10,0 % p/v); (e) adulteração com formol (10,0 % p/v); (f) adulteração com sacarose (10,0 % p/v). A região espectral utilizada para a construção dos modelos está delimitada pelas linhas tracejadas

A região entre 2000 e 900 cm^{-1} apresentou as maiores variações espectrais e foi considerada de melhor poder discriminante, contendo bandas características para quase todos os adulterantes. Então modelos locais foram testados e apresentaram erros de classificação menores quando comparados aos modelos que usaram os espectros inteiros. Por isso, os modelos finais foram desenvolvidos e validados utilizando apenas esta região espectral.

5.4.2 Modelos PLS-DA

Cinco modelos PLS1-DA foram construídos, um para cada adulterante. Cada modelo foi construído utilizando 155 amostras (15 amostras brancas, 100

amostras contendo apenas um adulterante, sendo 20 para cada um deles, e 40 amostras contendo o adulterante escolhido para aquele modelo em misturas com outros adulterantes). As amostras foram divididas em conjuntos de treinamento e de teste; 25% dessas amostras em cada modelo (38) foram selecionadas para o conjunto teste, sendo escolhidas em intervalos regulares de acordo com a composição das misturas. Para selecionar o melhor número de VLs, foi feita uma validação cruzada usando o critério das venezianas (*venetian blinds*), com 6 divisões. Como o conjunto de calibração é formado por um grande número de amostras e foi organizado agrupando-as sequencialmente em função da classe, esse critério é o mais apropriado para garantir uma seleção representativa e homogênea das amostras de todas as classes. Foi selecionado o número de VLs que apresentou o menor erro de classificação de validação cruzada (ECVC). A Tabela 6 mostra os valores de ECVC (a razão entre amostras erroneamente classificadas e número total de amostras) para cada modelo, assim como a variância explicada, que foi de 99,5% em **X** para todos os modelos, e variou entre 69,3 e 77,0% em **Y**. Os gráficos de classificação, com os valores preditos para todas as amostras podem ser vistos na Figura 33. Nestes gráficos, os dados estão ordenados de maneira a sempre começar pelas amostras que não contêm o adulterante de cada modelo. O LC para cada classe/modelo foi estimado através da teoria de Bayes, para minimizar os erros de classificação.⁷⁴

Tabela 6– Número de VLs, variâncias explicadas e erros de classificação de validação cruzada (ECVC) para os modelos PLS-DA desenvolvidos

Adulterantes	Variáveis Latentes	Variância Explicada - X (%)	Variância Explicada- Y (%)	ECVC
Água	9	99,5	76,2	0,07
Amido	8	99,5	72,5	0,07
Citrato de Sódio	9	99,5	77,0	0,02
Formol	9	99,5	69,3	0,04
Sacarose	9	99,5	73,1	0,07

Em todos os modelos, a primeira derivada com alisamento Savitzky-Golay foi utilizada, e os dados foram então centrados na média. Outros pré-

processamentos foram testados, como OSC, MSC e SNV, porém todos resultaram em ECVC mais elevados.

A interpretação espectral de modelos PLS ou PLS-DA frequentemente é feita a partir da observação dos seus coeficientes de regressão. Entretanto, como ressaltado por Brown & Green (2009),¹¹⁶ esta interpretação não deve se basear somente nos coeficientes de regressão dos modelos, uma vez que eles são dependentes da composição das amostras no conjunto de calibração/treinamento, da covariância implícita entre os componentes dessas amostras e da relação sinal/ruído dos dados.

Uma ferramenta mais adequada para a interpretação espectral de modelos PLS são os gráficos da Importância das Variáveis na Projeção dos escores (*Variable Importance in Projection, VIP scores*), que estimam a importância de cada variável na projeção utilizada pelo modelo PLS através dos coeficientes em cada componente, juntamente com a significância de cada componente na regressão.¹¹⁷ A Figura 34 mostra os VIP escores para cada modelo. É possível verificar uma alta concordância entre as atribuições espectrais apresentadas na seção 4.4.1 e as variáveis de maior importância observadas nos VIP escores.



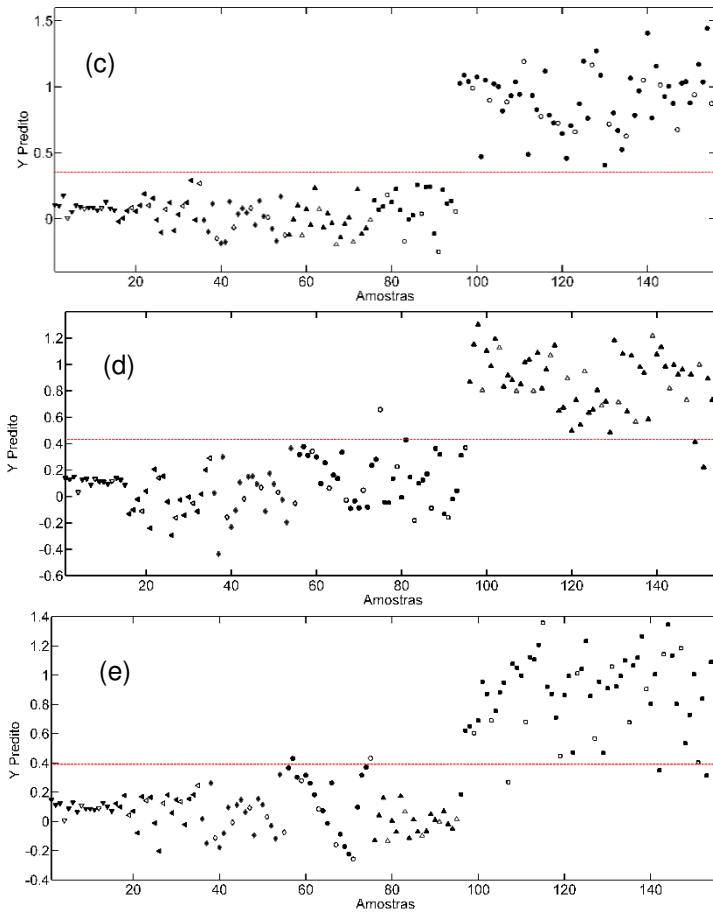
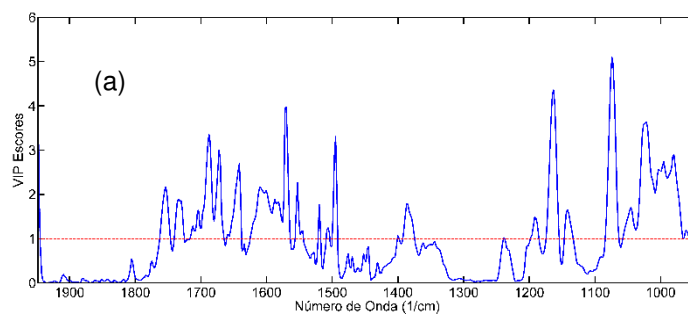


Figura 33 – Gráficos de previsão para adulterações com (a) água; (b) amido; (c) citrato de sódio; (d) formol; (e) sacarose. ∇ - Amostras Brancas; \triangleleft - adulteração com água; \diamond - Adulteração com amido; \circ - Adulteração com citrato; \triangle - Adulteração com formol; \square - Adulteração com sacarose. Marcadores cheios – Amostras de treinamento; Marcadores vazios – Amostras teste



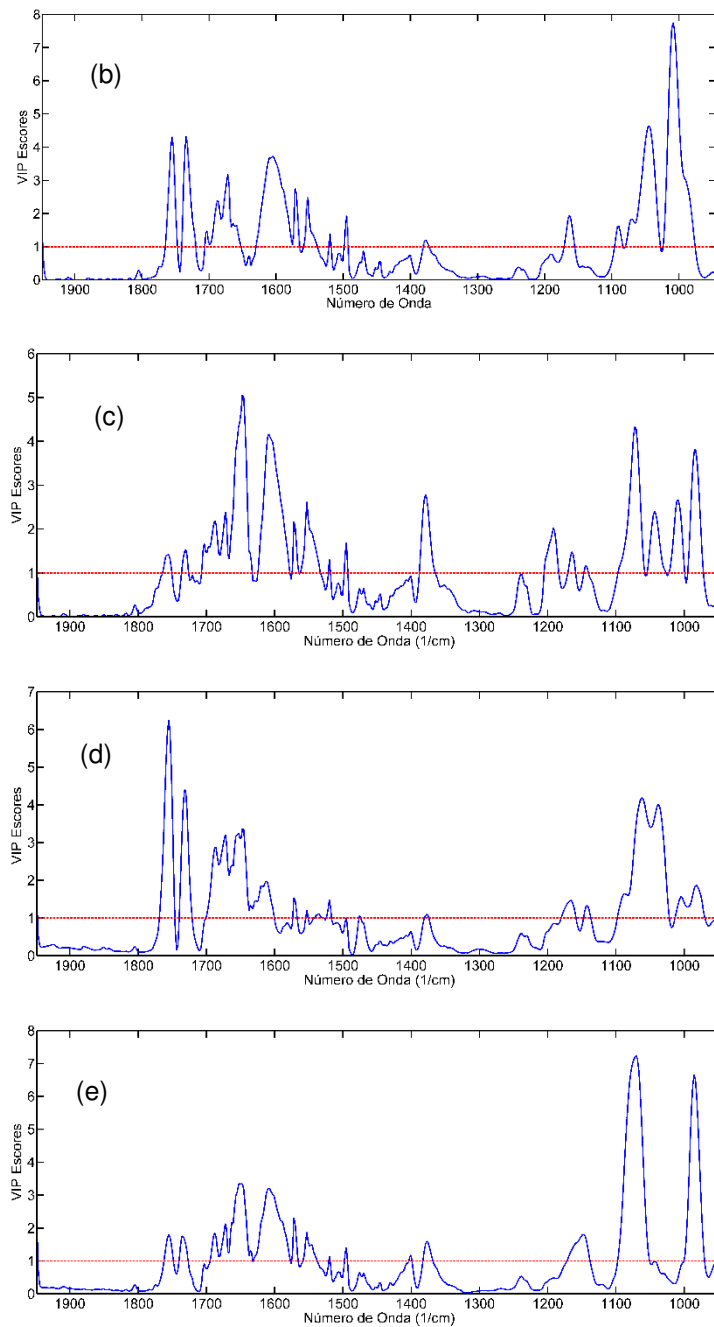


Figura 34 – VIP escores para os modelos de detecção de adulteração de leite cru com (a) água; (b) amido; (c) citrato; (d) formol; (e) sacarose

Como pode ser visto na Figura 34, todos os VIPs escores apresentaram picos distinguíveis em regiões especificamente características de cada adulterante, como os dois picos entre 1080 e 1040 cm^{-1} e 1015 e 980 cm^{-1} , característicos do citrato, uma banda larga entre 1100 e 1000 cm^{-1} , característica do formol, e os dois picos intensos em 1080 e 980 cm^{-1} , relativos à sacarose.

4.4.3 Validação qualitativa multivariada

Conforme mencionado anteriormente, a validação de métodos analíticos quimiométricos é um tema em desenvolvimento, não muito bem estabelecido, e a estimativa de FOM para método multivariados qualitativos tem recebido ainda menos atenção. Como é possível verificar na Tabela 7, sete diferentes FOM foram estimadas para os conjuntos de treinamento e teste.

O modelo para detectar adulterações com água apresentou os mais altos valores de TFP (4,0 %) e a segunda maior TFN (3,2 %) no conjunto de treinamento, além da maior TFP no conjunto de teste (11,5%). O número relativamente alto de classificações erradas neste modelo pode estar relacionado com a similaridade entre as amostras adulteradas com água e as amostras de branco (leite cru), e também com a detecção de amostras com níveis mais baixos de outros adulterantes (Figura 32).

O modelo de adulteração por amido apresentou baixas TFP e TFN no conjunto teste (3,0 e 4,8 %, respectivamente), com a classificação errônea de amostras contendo citrato e água sendo a ocorrência mais frequente. Este modelo também apresentou TFP e TFN relativamente elevadas no conjunto de teste (8,0 e 16,7 %, respectivamente). Os modelos para formol e sacarose apresentaram TFP (0,0 e 1,0 %, respectivamente) e TFN (3,2 e 4,8 %, respectivamente) similares no conjunto teste, e resultados idênticos no conjunto teste para a TFP (4,2 %) e a TFN (0,0 %). O modelo para detecção de citrato apresentou os melhores resultados, não produzindo nenhuma classificação equivocada nos conjuntos de treinamento e teste.

Tabela 7– FOM estimadas para os modelos de detecção de adulterantes em leite cru bovino

Conjunto Treinamento					
	TFP	TFN	TSL	TES	TCF
Água	4,0%	3,2%	96,8%	96,0%	92,8%
Amido	3,0%	4,8%	95,2%	97,0%	92,2%
Citrato de sódio	0,0%	0,0%	100,0%	100,0%	100,0%
Formol	0,0%	3,2%	96,8%	100%	96,8%

Sacarose	1,0%	4,8%	96,8%	99,0%	95,7%
Conjunto Teste					
	TFP	TFN	TSL	TES	TCF
Água	11,5%	0,0%	93,8%	88,5%	82,8%
Amido	8,0%	16,7%	83,3%	92,0%	75,3%
Citrato de Sódio	0,0%	0,0%	100,0%	100,0%	100,0%
Formol	4,2%	0,0%	93,8%	95,8%	89,6%
Sacarose	4,2%	0,0%	93,8%	95,8%	89,6%

TFP – Taxa de falso positivo; TFN – Taxa de falso negativo; TSL – Taxa de seletividade; TES – Taxa de especificidade; TCF – Taxa de confiabilidade

De acordo com a Diretiva 657/2002⁶⁷ da Comunidade Europeia, a TFN deve ser inferior a 5%. Resultados falso negativos devem ser mais rigorosamente controlados para garantir a segurança do consumidor, evitando que amostras adulteradas sejam classificadas como não adulteradas. Resultados falso positivos são menos preocupantes, pois todas as amostras positivas em um método de triagem devem ser confirmadas por um método de referência. Com exceção do modelo para a detecção de amido, todos os modelos propostos estão de acordo com o guia de validação acima citado.

Uma FOM mais representativa da veracidade, TCF, também consta na Tabela 7. Os valores de TCF para o conjunto de treinamento variaram entre 92,2 e 100 %, e para o conjunto teste variaram de 75,3 a 100 %, com os dois piores resultados sendo apresentados pelo modelo de detecção de amido.

A ACO foi estimada utilizando três replicatas de três diferentes níveis (0,5, 5,0 e 10 % p/v), analisadas em condições de repetitividade. A CON foi estimada utilizando o mesmo protocolo para a estimativa da CON, repetido em três dias diferentes por três analistas diferentes. O modelo para a adulteração de água apresentou os menores valores de ACO e CON (78 e 92 %, respectivamente), com todos os outros modelos apresentando ACO e CON de 100 %. Estes resultados sugerem que a detecção da adição de água no leite pelo método proposto não foi tão precisa, possivelmente devido ao elevado teor de água presente no leite naturalmente em amostras não adulteradas e à variação natural de seu teor.

4.5 Conclusão

Um método de triagem multivariado simples para a detecção simultânea de até cinco adulterantes em leite bovino cru foi desenvolvido e validado. O método foi capaz de detectar a presença específica de água, amido, citrato de sódio, formol e sacarose em amostras de leite cru em concentrações de até 0,5 % p/v. Todos os modelos, exceto o modelo para a detecção de amido, forneceram resultados consonantes com os guias de validação internacionais^{67,70}. O método proposto consome uma quantidade muito pequena de amostra (30 µL), é muito rápido (no máximo um minuto para aquisição de cada espectro), não destrutivo, não demanda nenhum pré-tratamento da amostra nem utiliza reagentes, além de analisar todos os adulterantes simultaneamente a partir dos mesmos espectros. Esta aplicação significou ainda um interessante exemplo de validação analítica de um método qualitativo multivariado, baseado em PLS-DA. Este assunto, raramente abordado na literatura, demanda uma maior atenção devido à necessidade de harmonização de conceitos, com vistas a uma futura incorporação do pensamento multivariado nas regulamentações oficiais sobre métodos qualitativos.

5. Determinação do corante amarelo crepúsculo em bebidas não alcoólicas utilizando análise multivariada de imagens

5.1 Introdução

Corantes industriais estão constantemente presente no estilo de vida moderno, sendo amplamente utilizados em cosméticos, roupas, fármacos e,

particularmente, em alimentos. Eles possuem uma série de vantagens se comparados aos corantes naturais, como maior estabilidade ao oxigênio, luz e mudanças de pH, boa solubilidade em água e baixo custo de produção.^{118,119} Os corantes azo constituem o maior grupo de corantes sintéticos, correspondendo a aproximadamente 70% de todos os corante artificiais conhecidos, e sua estrutura molecular se caracteriza pela presença de um grupo azo (-N=N-) localizado entre anéis aromáticos.¹¹⁹

Apesar de fornecerem diversos benefícios tecnológicos, e melhorarem as características sensoriais de um alimento, um grande número de estudos tem demonstrado efeitos colaterais provocados pelo consumo dos corantes azo, como respostas alérgicas, asma, urticária e imunossupressão, especialmente se consumidos em excesso.¹¹⁹

O Amarelo Crepúsculo (AC), também conhecido como *Sunset Yellow*, Amarelo Entardecer, E110 ou Amarelo Comestível 3 (Figura 35), é um dos corantes azo mais utilizados. Ele possui uma coloração alaranjada e é utilizado principalmente em produtos com sabor de frutas, como refrigerantes, sucos, doces e sorvetes. O AC também possui uma ampla utilização nas indústrias farmacêuticas e de cosméticos. Entretanto, o seu consumo pode causar alguns efeitos colaterais em humanos, e a sua ingestão tem sido associada a riscos de falência renal e danos hepatocelulares.¹¹⁹

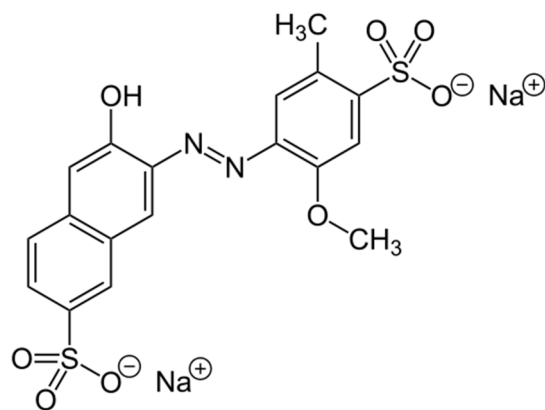


Figura 35– Estrutura química do Amarelo Crepúsculo

O grande aumento na utilização e consumo de corantes artificiais, principalmente em produtos destinados a crianças, criam uma necessidade de métodos que possam monitorar e quantificar estes corantes. No Brasil, a ANVISA (Agência Nacional de Vigilância Sanitária) é o órgão governamental responsável pela regulação de alimentos, e estabelece os limites para utilização de corantes artificiais em diferentes produtos. De acordo com a resolução R05/07, o limite para a concentração de amarelo crepúsculo em bebidas não alcoólicas é de 100 mg.L⁻¹.¹²⁰ Um dos métodos mais tradicionais para a determinação de corantes azo é baseado em espectrofotometria UV/VIS, o qual demanda extrações sequenciais com metanol contendo 5% de hidróxido de amônio.¹⁰⁷

Outros métodos, envolvendo diferentes técnicas analíticas, como cromatografia, potenciometria, voltametria, imunoensaios e extração no ponto nuvem com determinação espectrofotométrica foram descritos na literatura^{118,121–124}. Estratégias quimiométricas também têm sido aplicadas na determinação de corantes em alimentos, principalmente usando espectrofotometria UV/Vis, avaliando misturas binárias ou ternárias de corantes.^{125–127}

5.2 Objetivos

Desenvolver e validar um método analítico multivariado simples e rápido utilizando imagens digitais obtidas em um escâner de mesa comercial para a determinação do corante amarelo crepúsculo em bebidas não alcoólicas (refrigerantes e isotônicos).

5.3 Materiais e métodos

5.3.1 Equipamentos e softwares

As imagens foram obtidas utilizando um escâner de mesa comercial CanoScan LiDE 110 (Tokio, Japão). Os *softwares* utilizados foram os mesmos citados na sessão 3.3.1, juntamente com o *Image Processing Toolbox*, versão 8.0 (The MathWorks, Natick, MA, EUA).

5.3.2 Amostras

Oitenta e três amostras de bebidas não alcoólicas de vinte e cinco marcas diferentes (refrigerantes de laranja e bebidas isotônicas de laranja e tangerina) contendo o AC como único corante artificial foram adquiridas predominantemente no comércio local de Belo Horizonte/MG, mas também em diversos outros locais do Brasil, e armazenadas a 4 °C até a análise.

5.3.3 Procedimento

As amostras foram deixadas por 30 minutos em descanso sobre a bancada do laboratório, para atingirem o equilíbrio térmico antes do início dos procedimentos. 30 mL de cada amostra foram coletados em um béquer de 50 mL e levados ao banho de ultrassom, para degaseificação, por 5 minutos. Após esta etapa, 1 mL da amostra foi destinado à análise cromatográfica, para a obtenção dos valores de referência, e 5 mL foram utilizados para a aquisição das imagens.

A aquisição das imagens foi realizada utilizando placas de Petri pequenas (5,0 cm de diâmetro x 1,5 cm de altura), nas quais as amostras a serem lidas eram vertidas, e as placas posicionadas no canto inferior esquerdo do escâner. Um anteparo foi utilizado para evitar a entrada de luz externa (Figuras 36 e 37). Todas as imagens foram digitalizadas no sistema RGB 24-bit, com 16,8 milhões de cores e resolução de 300 ppp (pixels por polegadas), no formato “.tif”. A conversão das imagens em histogramas RGB foi realizada em ambiente MatLab. Primeiramente, uma área de 100x100 pixels de cada imagem foi selecionada de uma área central da placa de Petri, em uma parte homogênea da imagem. A área selecionada foi tratada com um filtro digital (*unsharp*) para redução de ruído, e então decomposta em histogramas RGB. Após todos os tratamentos, um

histograma contendo a frequência de 768 canais (256 para cada cor RGB) foi obtido para cada amostra (Figura 38). Cada amostra foi escaneada três vezes, e a média dos três histogramas de cada amostra foi utilizada para a construção do modelo PLS.



Figura 36 – Escâner CanoScan LiDE 110 com o anteparo

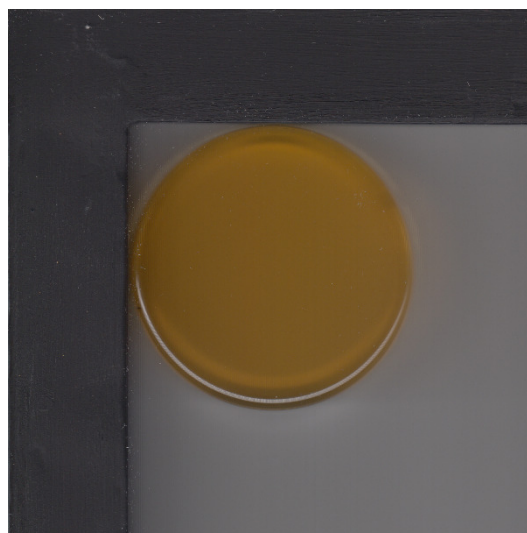


Figura 37 – Imagem adquirida de uma amostra de refrigerante

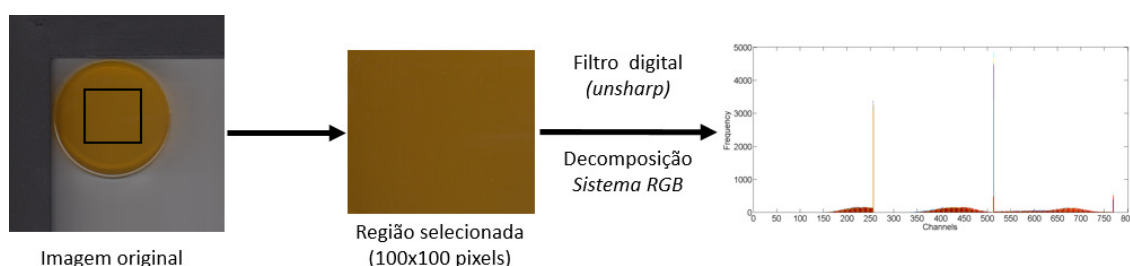


Figura 38– Representação esquemática da decomposição matemática das imagens

5.3.4 Análises Cromatográficas

A metodologia utilizada para a obtenção dos valores de referência das amostras foi adaptada a partir de um método cromatográfico desenvolvido por um fabricante de equipamentos.¹²⁸ As análises foram realizadas em equipamento de cromatografia líquida de alta eficiência Finningan Surveyor HPLC system (Thermo Fisher Scientific, San Jose, EUA) com detecção por arranjo de diodos (CLAE-DAD), utilizando uma coluna C-18 Shimadzu Shim-Pack XR-ODS (3 mm d.i. x 150 mm comp.). Foi utilizada eluição por gradiente, com a fase móvel sendo composta por uma solução aquosa de acetato de amônio 20 mM e uma fase orgânica de acetonitrila/metanol (80:20, v/v). Um fluxo de $1,2 \text{ mL}\cdot\text{min}^{-1}$ foi utilizado e a detecção foi realizada em 484 nm. O tempo total da corrida foi de 15 minutos, com o tempo de retenção do AC de 7,5 min.

5.4 Resultados e discussão

5.4.1 Modelo PLS

Todas as amostras analisadas continham apenas AC como o único corante artificial adicionado. Previamente à construção do modelo multivariado utilizando os valores de referência obtidos por CLAE para as amostras reais, uma estratégia alternativa foi testada. Um modelo multivariado construído com dados de imagens de soluções padrão de AC foi utilizado na predição das amostras reais, mas não apresentou bons resultados. Todas as bebidas analisadas continham aditivos e outros ingredientes, o que tornou as amostras um pouco turvas, impedindo que as mesmas fossem preditas com erros satisfatórios utilizando um modelo obtido a partir de soluções padrão. Além disso, diversas amostras continham uma pequena quantidade de suco natural, e conseqüentemente, corantes naturais, interferentes que contribuem diretamente para a cor, justificando mais uma vez a utilização de um modelo multivariado construído a partir de amostras reais. Dessa forma, um modelo PLS foi construído utilizando as imagens de refrigerantes e bebidas isotônicas e os valores de referência (concentração de AC) foram obtidos através de um método cromatográfico.

A Figura 39 mostra os histogramas RGB para as 83 amostras analisadas. É possível notar que o canal 513 é o de maior frequência, seguido pelo canal 256, o que indica uma predominância de tons amarelados nas imagens (mistura das cores vermelha e verde). Como esses dois canais apresentaram maiores frequências nos histogramas, modelos univariados para cada uma destas duas frequências foram construídos. Estes modelos foram comparados aos modelos multivariados através de seus coeficientes de correlação (r) entre os valores de referência e os valores preditos. Boas correlações foram encontradas (0,827 para o canal 513 e -0,778 para o canal 256), mas estes valores não foram considerados satisfatórios para um modelo de quantificação. A remoção de algumas áreas do histograma também foi avaliada, porém não foi observada melhora nos modelos.

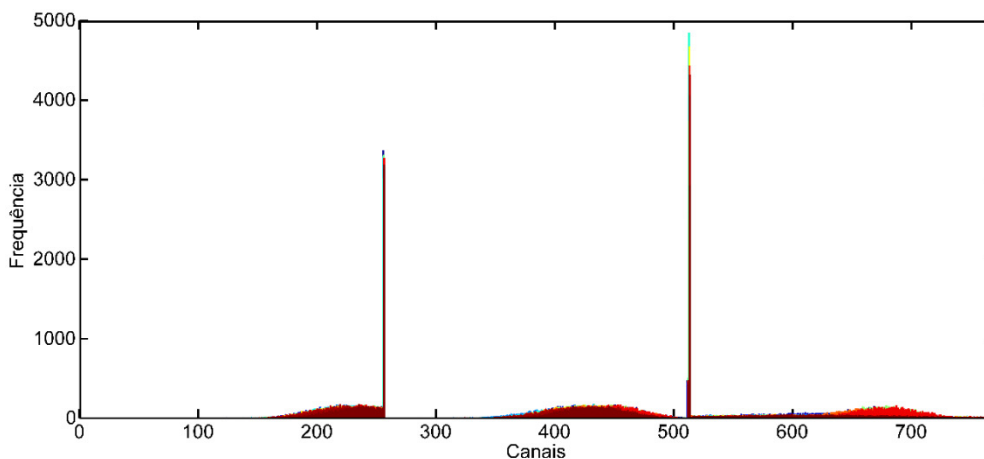


Figura 39– Histograma de frequência para as amostras de bebidas não alcoólicas analisadas

Para a construção do modelo, os histogramas das amostras foram divididos em dois terços para o conjunto de calibração (56 amostras) e um terço para o conjunto de validação (27 amostras), utilizando o algoritmo de Kennard-Stone.⁹³ Os histogramas foram apenas centrados na média, sem nenhum outro pré-processamento. O número de variáveis latentes foi selecionado com base no menor valor de RMSECV, estimado através da validação cruzada do tipo veneziana (6 divisões), sendo o melhor modelo construído com 4 VLs. Os coeficientes de regressão estimados para o modelo PLS são mostrados na Figura 40.

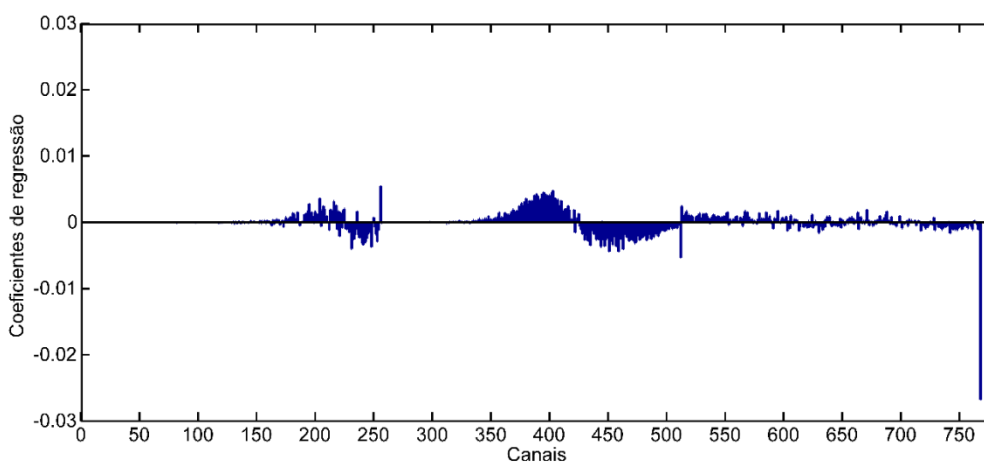


Figura 40– Coeficientes de regressão do modelo PLS para determinação de AC em bebidas não alcoólicas

Através da observação dos coeficientes na Figura 40, é possível identificar os canais que mais contribuem para a previsão do AC. O coeficiente

mais significativo foi observado no canal 768, que mostrou uma influência negativa para quantificação do AC, indicando uma relação inversamente proporcional entre tons azuis e a concentração de AC. O maior coeficiente positivo foi apresentado pelo canal 256, porém, o coeficiente de regressão do canal 513, o canal mais frequente nos histogramas das amostras, não foi muito significativo, o que indica que os sinais mais intensos não são, necessariamente, os mais importantes para a predição. Um modelo univariados correlacionando as frequências no canal 768 e os valores de referência foi construído, porém, sem apresentar bons resultados ($r=0,539$).

O modelo PLS foi otimizado através da detecção de *outliers* e avaliado com base no RMSEP e no RPD, como é mostrado na Tabela 8. Onze *outliers* foram removidos do conjunto de calibração (19,6%), nove por em alto resíduos em Y e dois por alto *leverage*. Cinco *outliers* (18,5%) foram removidos do conjunto de validação, todos por alto resíduo em Y. O modelo final, com 4 VLs, explicou 98,8% da variância em X e 97,2% em Y.

Tabela 8 - Resultados da otimização do modelo PLS através da detecção de *outliers* (modelo final em negrito)

Modelo	1 ^o	2 ^o	3 ^o	4 ^o
Amostras de calibração	56	51	48	45
Amostras de validação	27	27	27	22
Variáveis latentes	4	4	4	4
RMSEC (mg L ⁻¹)	2,7	2,1	1,7	1,3
RMSEP (mg L ⁻¹)	2,6	2,9	3,1	1,3
RPD calibração	2,3	3,1	3,8	4,6
RPD validação	3,3	3,0	2,8	6,5

5.4.2 Validação Analítica Multivariada

A Tabela 9 apresenta as FOM estimadas para o modelo otimizado. O gráfico dos valores preditos vs. valores de referência é mostrado na Figura 41. Para confirmar a linearidade, testes estatísticos apropriados foram utilizados para avaliar a linearidade (RJ), homocedasticidade (BF) e ausência de correlação dos resíduos (DW), a 95% de confiança. Os resultados dos testes confirmam a linearidade do modelo, principalmente o DW (1,86), que apresentou um valor dentro dos limites de aceitação (1,5 – 2,5), confirmando a independência dos resíduos. Uma vez considerado linear, os parâmetros para a

regressão linear foram estimados, incluindo o coeficiente de correlação entre valores preditos e valores de referência (r) de 0,9742.

A veracidade foi estimada através de parâmetros que avaliam os erros absolutos, como o RMSEP ($1,3 \text{ mg L}^{-1}$). Erros relativos individuais de predição para cada amostra também foram estimados, apresentando valores entre -6,2 e 9,0 %, com valor médio de -2,0% que estão de acordo com o guia de validação da AOAC⁴¹ (-20/+10). Os valores de precisão sob condições de repetitividade e precisão intermediária, expressos como desvio padrão relativo (DPR_r e DPR_R , respectivamente), foram estimadas por análise de variância dos valores preditos para 18 replicatas de amostras em três níveis de concentração distintos.⁴⁷ Os valores de DPR_R variaram entre 4,1 e 5,6%, e os valores de DPR_r variaram entre 1,0 e 3,1%. Todos os valores estimados estão abaixo do limite máximo, calculado através da equação de Horwitz.⁴⁰

Os resultados de veracidade e precisão corroboram com a afirmação de que o método pode ser considerado exato. Considerando a exatidão e a linearidade, a faixa de trabalho do método foi definida como sendo de 7,8 a 39,7 mg L^{-1} de AC.

Tabela 9– Parâmetros para avaliação das FOM no modelo proposto para a determinação de AC em bebidas não alcoólicas

Figuras de Mérito	Parâmetro	Valores
Veracidade (n=22)	Erro relativo médio	-2,0 %
	DPR repetitividade ^a (DPR_r)	2,9 % (7,2 %)
		1,0 % (6,2 %)
Precisão (n=18)	DPR precisão intermediária ^b (DPR_R)	3,1 % (6,4%)
		4,4 % (11,0 %)
		5,6 % (9,2 %)
Linearidade	Durbin-Watson	1,86
	Inclinação ^c	$0,97 \pm 0,02$
	Intercepto ^c	$0,80 \pm 0,74$
	Coeficiente de correlação	0,9702
Faixa de trabalho		7,8 - 39,7 mg L^{-1}
Seletividade do modelo		0,17

Sensibilidade do modelo	154,11 ^d
Sensibilidade analítica (γ)	3,4 L mg ⁻¹
γ^1	0,3 mg L ⁻¹
Viés	0,512±1,210 mg L ⁻¹

^a Resultados para três amostras em três diferentes concentrações analisadas no mesmo dia por um mesmo analista, em sextuplicata. Valores limites de DPR_r (entre parênteses) calculados através da equação de Horwitz para cada nível de concentração estudado.

^b Resultados para três amostras em três diferentes concentrações em dias diferentes por analistas diferentes, em sextuplicata. Valores limites de DPR_R (entre parênteses) calculados através da equação de Horwitz para cada nível de concentração estudado.

^c Parâmetros do ajuste para a curva de valores de referência vs. valores preditos para as amostras de calibração

^d Valores expressos como a razão entre frequência e mg L⁻¹.

Na Figura 41 é possível notar que as amostras não se distribuem de maneira homogênea ao longo da curva, formando 3 grupos claramente distinguíveis. O primeiro grupo, com as menores concentrações, é formado pela amostras de isotônicos. O segundo grupo, no meio da curva, é formado pelas amostras mais famosas de refrigerantes de laranja, e o terceiro grupo é formado por refrigerantes de laranja menos conhecidos e de menor preço no mercado, que apresentavam as maiores concentrações de AC.

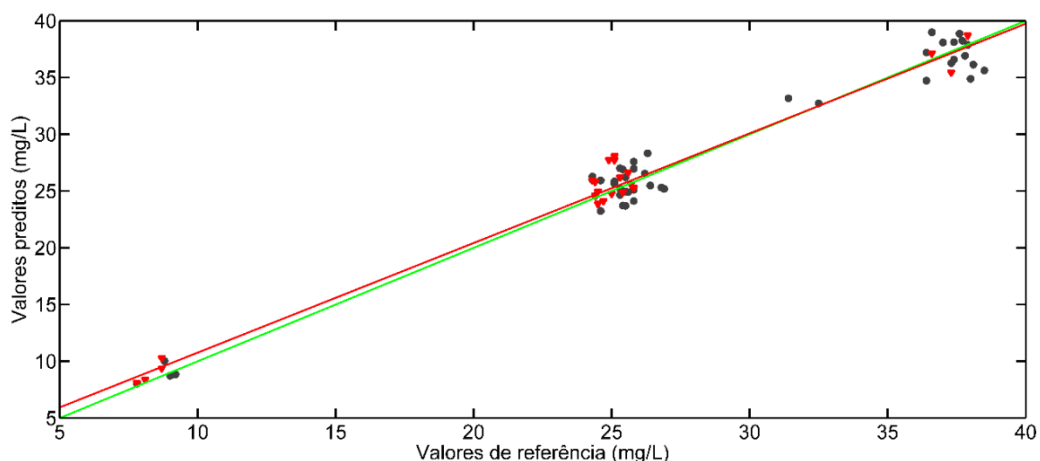


Figura 41– Gráfico de valores de referência vs. valores preditos. Calibração (quadrados) e validação (triângulo)

A SEN e a SEL também foram estimadas, com base no NAS. A necessidade de um valor mínimo de SEL não tem sentido prático na validação multivariada, uma vez que, diferente da validação univariada, métodos multivariados podem ser acurados mesmo com valores de SEL baixos. A SEL estimada indica que 17% do sinal analítico original foi utilizado no modelo para

a determinação de AC. A γ foi estimada através do ruído instrumental (44,7), obtido da medida de dez replicatas de imagens de uma placa de Petri vazia. O seu inverso (γ^{-1}), 0,3 mg.L⁻¹, indica a quantidade mínima discernível pelo método, considerando o erro instrumental aleatório como única fonte de erro, e também define o número de casas com o qual os resultados devem ser expressos.

O viés, estimado apenas para o conjunto de calibração, apresenta um valor de t (1,98) menor que o valor crítico (2,06, com 22 graus de liberdade e 95% de confiança), o que indica a ausência de erro sistemático nas predições do modelo. De acordo com a literatura,^{56,57} bons modelos de calibração devem apresentar valores de RPD acima de 2,4, enquanto valores entre 2,4 e 1,5 são considerados aceitáveis. Considerando os valores apresentados na Tabela 9, as estimativas de RPD foram satisfatórias tanto para a calibração quanto para a validação.

O EAEP, calculado com a equação EIV (*Errors in variables*), variou entre 3,3 e 3,6 mg.L⁻¹. A Figura 42 mostra os valores de referência e os valores preditos com os respectivos intervalos de confiança para cada amostra. Para essa estimativa, a incerteza do método de referência (0,03 mg.L⁻¹) foi obtida através de estudos de precisão intermediária. Cinco pseudo-graus de liberdade⁵⁹ foram utilizados na estimativa dos intervalos de confiança ($t=2,77$). Para todas as amostras, os valores de referência estão dentro do intervalo de confiança estimado.

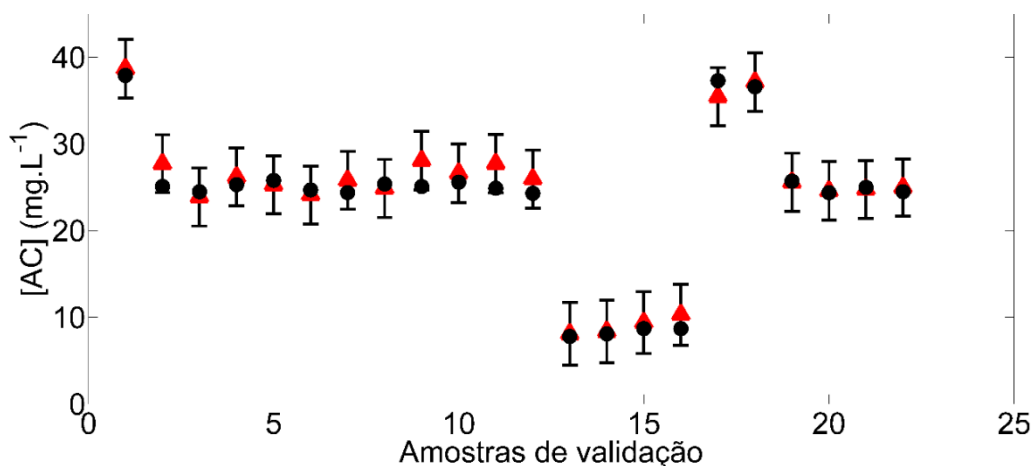


Figura 42– Intervalos de confiança para a predição das amostras de validação, estimados com base no EAEP. Valores preditos (triângulos) e de referência (círculos)

5.4.3 β -TI

O β -TI estimado para o modelo pode ser visto na Figura 43. Este perfil de acurácia mostra que os limites de confiança calculados estão contidos no intervalo definido pela AOAC.⁴¹ Isto indica que é esperado que todos os valores obtidos utilizando este método irão apresentar erros relativos dentro do limites da legislação. Para a menor concentração avaliada (9,6 mg.L⁻¹), o β -TI variou entre -15,4 e 9,6 %. Para a concentração intermediária (30,0 mg.L⁻¹), o β -TI variou entre -13,7 e 6,0 %, e para concentração mais elevada (38,8 mg.L⁻¹), ele variou entre -12,2 e 9,4 %.

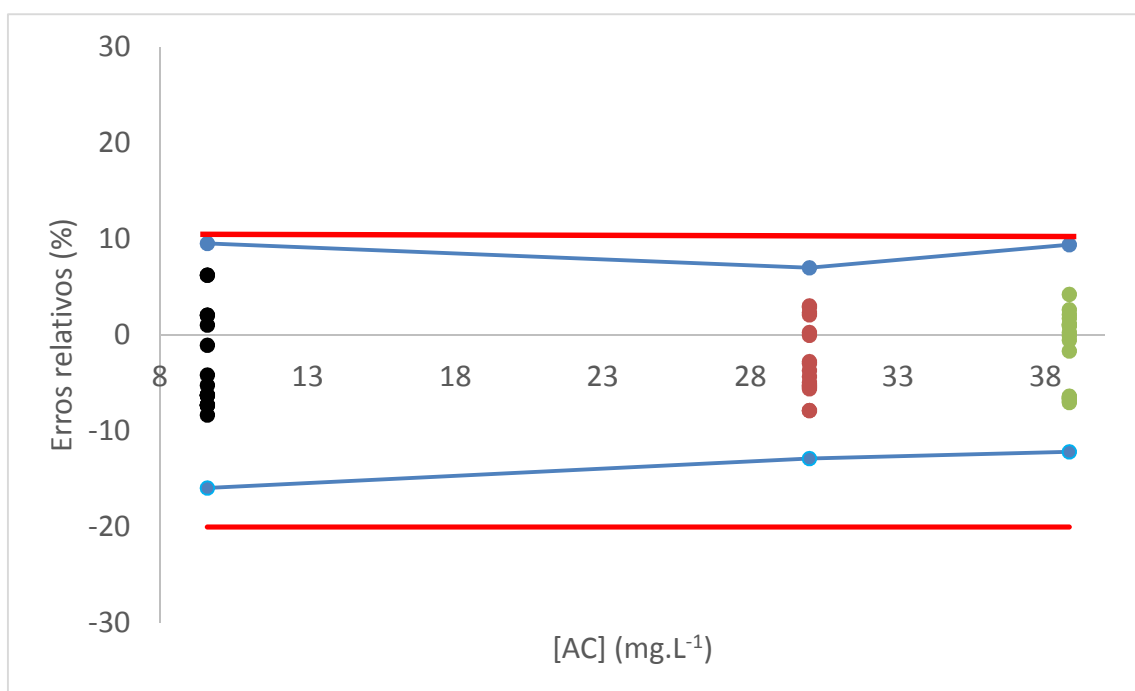


Figura 43 - β -TI estimados para o modelo (linhas azuis) e erros máximos relativos definidos pelo guia de validação brasileiro (linhas vermelhas)

5.5 Conclusões

Um método de calibração multivariada simples baseado em histogramas RGB obtidos a partir de imagens digitais foi desenvolvido e validado para a determinação de amarelo crepúsculo em bebidas não alcoólicas. Este método utilizou um equipamento de baixo custo (escâner de mesa comercial), não necessitou de pré-tratamento das amostras ou utilizou solventes e reagentes, além de ter sido muito mais rápido que o método de referência (poucos segundos contra 15 minutos da corrida cromatográfica). Adicionalmente, o método proposto pode ser utilizado para automatização *online* de processos industriais com uma frequência de amostragem elevada, ou em equipamentos portáteis. O método foi completamente validado de acordo com guias brasileiros e internacionais de validação, sendo considerado linear, acurado, com ausência de viés e adequado para a utilização como metodologia oficial para quantificação de corantes artificiais em alimentos.

6. Telefone celular como instrumento de aquisição de imagens para modelos multivariados: Determinação do corante vermelho 40 em balas duras

As balas são produtos obtidos a partir do cozimento de açúcares, adicionados de corantes, aromatizantes e acidificantes, podendo conter ainda outras substâncias permitidas, características para cada tipo de bala. Elas podem ser classificadas em balas duras e mastigáveis.¹²⁹

A bala dura é um produto preparado à base de açúcares e adicionado de substâncias, como corantes, ácidos e aromatizantes, devendo estar em estado vítreo, tendo características de translucidez, com textura dura e quebradiça, e com teor de sólidos solúveis de 97 a 98%. As balas mastigáveis se diferenciam pelo teor de sólidos solúveis, que varia de 90 a 93%, e pela adição de gorduras, que lhes fornecem maior consistência e maciez.¹²⁹

O Brasil é o terceiro maior produtor mundial de balas e confeitos, superado apenas pelos EUA e Alemanha, tendo exportado, em 2004, 153.000 toneladas para 144 países, gerando divisas de R\$ 161 bilhões.¹²⁹ De acordo com a legislação brasileira, denominam-se balas e caramelos as preparações à base de pasta de açúcar fundido, de formatos variados e de consistência dura ou semidura, com ou sem adição de outras substâncias permitidas.¹³⁰ Normalmente, são adicionados às balas acidulantes, aromatizantes, e corantes.

A utilização de corantes em alimentos desencadeia uma série de polêmicas, pois a principal justificativa para seu emprego é tornar o produto mais atrativo esteticamente. Somado a isto, diversos estudos vêm demonstrando a ocorrência de reações adversas a curto e longo prazo, associadas ao consumo de alimentos que apresentam esses aditivos. Portanto, já se sabe que eles não são totalmente inofensivos à saúde.¹³¹

O público infantil é o maior consumidor de alimentos coloridos, pois a indústria investe maciçamente nesse segmento visando as crianças e tornando seus produtos mais atrativos a esta faixa etária. Porém, a presença de reações alérgicas não é rara, pois as crianças apresentam maior suscetibilidade às

reações adversas provocadas pelos aditivos alimentares, devido à sua “imaturidade fisiológica”, que prejudica o metabolismo e a excreção dessas substâncias.¹³² Além disso, a capacidade cognitiva de um adulto para controlar o consumo regular de alimentos ainda não é observada em uma criança.¹³¹ O método mais usado para a determinação de corantes em balas é o mesmo citado no capítulo anterior, para a determinação de corantes em refrigerantes.¹⁰⁷ Ele é um método que consome uma grande quantidade de solvente (metanol com 5% m/v de NH₄OH) e é demorado, em função de requerer sucessivas extrações. Com isso, tem-se a necessidade de novos métodos também para a determinação do teor de corantes nesse tipo de produto. Métodos alternativos de análise de corantes em balas também foram propostas por diferentes autores, utilizando diferentes técnicas, como polarografia de pulso diferencial¹³³, cromatografia¹³⁴ e UV conjuntamente com métodos multivariados de ordem superior¹³⁵ (mínimos quadrados bilineares com bilinearização residual).

O contínuo progresso no aperfeiçoamento e aumento das capacidades de telefones celulares tem feito destes equipamentos mais que simples aparelhos de comunicação, transformando-os em verdadeiros computadores portáteis, equipados com câmeras cada vez mais potentes. A miniaturização de equipamentos e o conceito de *Lab on a chip* (laboratório em um chip) encontraram nos celulares modernos uma grande ferramenta, pois eles conciliam, no mesmo equipamento, alta capacidade de aquisição de dados e alto poder de processamento.¹³⁶ A utilização de telefones celulares em métodos analíticos é um assunto recente, que surgiu juntamente com a popularização dos mesmos. Atualmente, eles são utilizados nas mais diversas áreas, como avaliação de estágio de maturação de frutas^{137,138}, análise de água¹³⁹⁻¹⁴¹, diferenciação de peças de carne¹¹

O vermelho 40, também conhecido como Allura Red AC, Food Red 17, E-129 ou CI 16035 (Figura 44), é um dos corantes do tipo azo, que teve seu uso aprovado pelo FDA (*Food and Drug Administration*) em 1971. Este corante apresenta boa estabilidade à luz, calor e ácido, além de ser o corante vermelho mais estável para bebidas na presença do ácido ascórbico, um agente redutor. Diversos estudos metabólicos mostraram que o vermelho 40 é pouco absorvido

pelo organismo e em estudos de mutagenicidade ele não apresentou potencial carcinogênico.¹⁴²

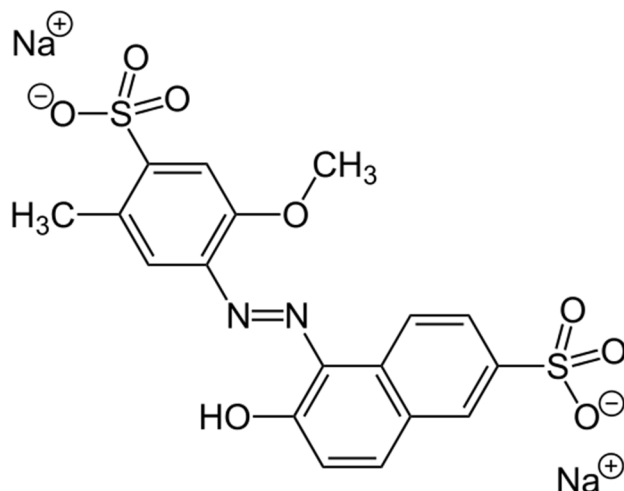


Figura 44 – Estrutura molecular do vermelho 40

Diferentemente das amostras de refrigerantes, as balas se encontram no estado sólido, e para a aquisição das imagens de maneira similar à realizada na aplicação anterior, elas teriam que ser solubilizadas em água. Esse procedimento aumentaria o tempo total de análise e poderia aumentar o erro analítico, em função da maior manipulação da amostra. Por isso, a alternativa foi a aquisição das imagens digitais dessas amostras no estado sólido. Dessa maneira, o processo da aquisição das imagens foi facilitado e a solubilização das balas não é necessária. Ao se trabalhar com as amostras no estado sólido, porém, um novo problema surge, que é a variação textural das amostras. Embora a cor das balas costume ser homogênea, sempre são observadas certas variações texturais em suas superfícies, as quais dificilmente são perfeitamente lisas. A decomposição das imagens no sistema RGB não consegue modelar variações texturais, pois cada pixel é decomposto individualmente, sem levar em conta a sua vizinhança.¹⁷

Uma possível solução para contornar o problema da variação textural é a utilização de outra estratégia de tratamento de imagens digitais que, ao invés de utilizar os histogramas RGB, trata diretamente os valores de intensidade de cor em função de sua posição (localização nos pixels x e y), gerando dados multidimensionais.¹⁵ Esses dados devem então ser pré-processados por uma

FFT para que se tornem congruentes e adquiram trilinearidade, condição para que sejam tratados por métodos quimiométricos de ordem superior, como por exemplo o NPLS.²² A descrição completa desta metodologia é apresentada no último parágrafo da seção 2.5 (Análise de imagens digitais) desta tese e também pode ser encontrada no artigo de Huang e coautores.¹⁷

6.2 Objetivos

Desenvolver um modelo N-PLS para determinação do corante vermelho 40 em balas duras de diferentes sabores (morango, framboesa, *tutti frutti* e cereja) utilizando imagens digitais obtidas com um telefone celular.

6.3 Materiais e métodos

6.3.1 Equipamentos e softwares

As imagens foram obtidas utilizando um celular da marca Motorola RAZR, modelo XT910 (Illinois, EUA), mostrado na Figura 45. Os *softwares* utilizados foram os mesmos citados na sessão 3.3.1, juntamente com o *Image Processing Toolbox*, versão 8.0 (The MathWorks, Natick, MA, EUA).



Figura 45 – Celular Motorola RAZR, empregado como instrumento para a aquisição das imagens

6.3.2 Amostras

Duzentos e quarenta amostras de balas duras de quatro sabores diferentes (morango, framboesa, cereja e *tutti frutti*) e de duas marcas distintas,

contendo o vermelho 40 como único corante artificial, foram adquiridas predominantemente no comércio local de Belo Horizonte/MG e armazenadas a temperatura ambiente até a data da análise.

6.3.3 Procedimento

As amostras foram desembaladas e as imagens foram imediatamente adquiridas. Para a aquisição das imagens foi construída uma câmara, com material reciclado (uma caixa de papelão). Esta câmara tem por objetivo padronizar a posição da aquisição das imagens, através de suportes para o celular e para a amostra (feito com durepox), assim como padronizar a iluminação, o que foi feito utilizando uma luminária alimentada via USB (Figuras 46-48). Após a aquisição das imagens, as balas foram solubilizadas em 30 mL de água deionizada aquecida (50 – 60 °C) por 15 minutos e analisadas pelo método de referência, uma adaptação do método oficial substituindo a solução metanólica de NH_4OH a 5% por água deionizada. As leituras foram realizadas em um equipamento Varian Cary100 UV/VIS, utilizando cubetas de quartzo de 10 mm e leitura em 505 nm.

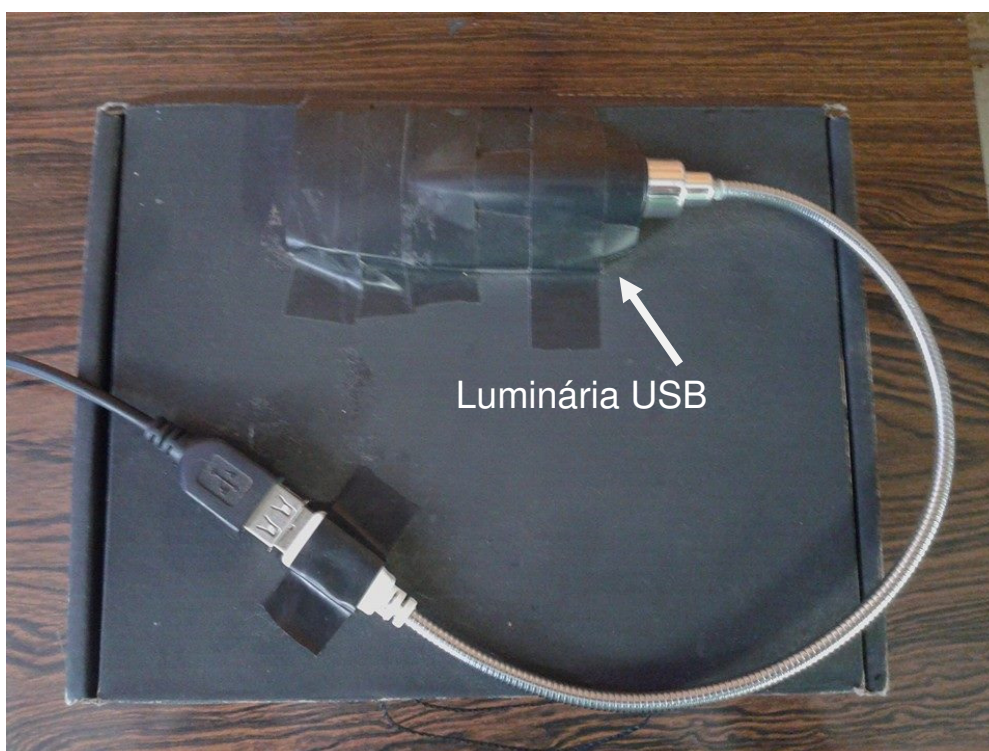


Figura 46 – Câmara de aquisição de imagens (vista superior fechada)



Figura 47 - Câmara de aquisição de imagens (vista superior aberta)



Figura 48 - Câmara de aquisição de imagens (vista lateral)

Para a aquisição das imagens, as balas foram posicionadas no suporte interno, o celular encaixado no suporte externo e a luminária USB ligada. As

fotos foram adquiridas sem *flash*, em triplicatas, em formato “.jpg” (Figura 49). A FFT foi aplicada nas imagens em ambiente MatLab. Primeiramente, uma região de 751x351 pixels foi selecionada, buscando abranger a maior área da amostra possível. Em seguida, a área selecionada foi convertida para escala de cinza, visando reduzir a complexidade dos dados e o tempo computacional gasto no cálculo dos modelos, e uma média foi estimada para as triplicatas das amostras. Após o tratamento dos dados, uma matriz 240 x 751 x 351 foi gerada.



Figura 47 – Imagem obtidas pelo celular de uma das amostras de bala analisadas

6.4 Resultados e discussão

6.4.1 Modelo PLS/RGB

Inicialmente, um modelo PLS utilizando imagens decompostas através do sistema RGB foi utilizado. Apesar de aceitáveis, os erros foram considerados elevados, com RMSEC de 6,4 mg.Kg⁻¹ e RMSEP de 9,2 mg.Kg⁻¹, após a remoção de *outliers*. Conforme exposto anteriormente, a decomposição em sistema RGB não modela muito bem a informação textural. Para superar este problema, a FFT foi testada.

6.4.2 Modelo N-PLS/FFT

Para a construção dos modelos, as amostras foram divididas em dois terços para o conjunto de calibração (160 amostras) e um terço para o conjunto de validação (80 amostras), utilizando o algoritmo de Kennard-Stone⁸⁹ aplicado aos seus histogramas RGB, e nenhum pré-processamento foi utilizado. O número de variáveis latentes do modelo N-PLS foi estimado com base no RMSECV, utilizando validação cruzada do tipo veneziana (7 divisões), sendo o melhor modelo obtido com 6 VLs.

O modelo N-PLS foi otimizado através da detecção de *outliers* e avaliado com base no RMSEP e no RPD, como é mostrado na Tabela 10. 36 *outliers* (22% do total) foram removidos do conjunto de calibração e 16 (21% do total) foram removidos do conjunto de validação. Neste capítulo em específico, a detecção de *outliers* foi realizada utilizando o teste *Jackknife*. Esta estratégia foi adotada devido ao elevado tempo computacional requerido para o cálculo dos modelos. Como o método de *Jackknife* remove todos os *outliers* de uma única vez, apenas o modelo inicial e o otimizado foram calculados. Destaca-se, portanto, que nesta aplicação a detecção de *outliers* se limitou aos resíduos de previsão (no bloco Y).

Tabela 10 - Resultados da otimização do modelo N-PLS através da detecção de *outliers* (modelo final em negrito)

Modelo	1 ^o	2 ^o
Amostras de calibração	160	124
Amostras de validação	78	62
Variáveis latentes	6	6
RMSEC (mg.Kg ⁻¹)	4,8	3,8
RMSEP (mg.Kg ⁻¹)	9,3	6,8
RPD calibração	3,2	4,1
RPD validação	1,7	2,3

6.4.3 Validação analítica multivariada

A Tabela 11 apresenta as FOM estimadas para o modelo otimizado, e o gráfico de valores preditos vs. valores de referência é apresentado na Figura 50. A linearidade do modelo foi confirmada através de testes estatísticos para avaliar a linearidade (RJ), homocedasticidade (BF) e a ausência de correlação dos

resíduos (DW), a 95% de confiança. Os resultados dos testes confirmam a linearidade, especialmente o valor de DW (1,96). Após a confirmação da linearidade, a regressão linear foi calculada e os seus parâmetros estão descritos na Tabela 11.

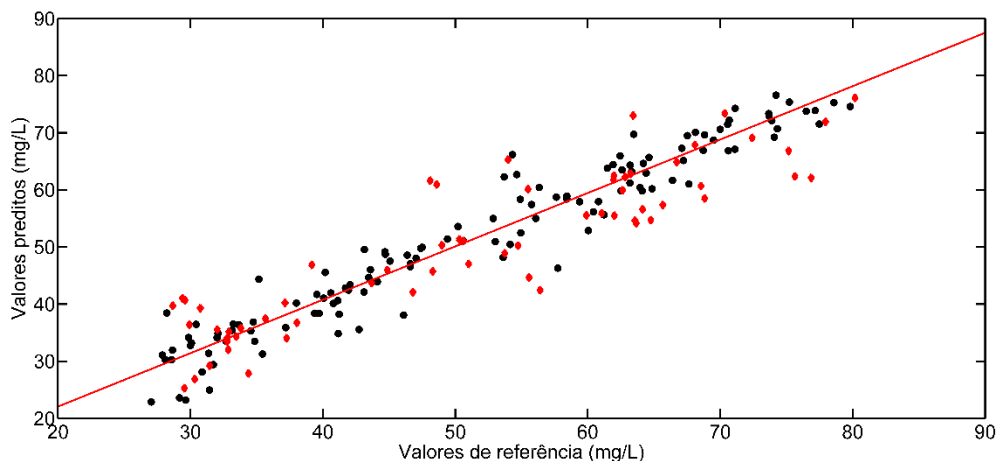


Figura 48 – Gráfico de valores preditos vs. valores de referência para o modelo de determinação do corante vermelho 40 em balas duras

Ao alterar o domínio dos dados, a FFT faz com que os coeficientes de regressão ou VIP Scores não sejam interpretáveis, por isso não houve discussão dos coeficientes de regressão para este capítulo em específico.¹⁷

A veracidade foi estimada através de parâmetros absolutos, como o RMSEP ($6,8 \text{ mg.Kg}^{-1}$), e através dos erros relativos de predição, que variaram entre 39,5 e -24,8%. Apesar de alguns poucos erros elevados, a grande maioria das amostras (80% do total) apresentou erros dentro da faixa aceitável, que varia entre -20%/+10% (Figura 51).⁴¹ A sensibilidade, sensibilidade analítica e γ^1 foram estimados através do NAS, calculado utilizando a equação MKL¹⁴³, e o ruído instrumental calculado através do desdobramento da matriz das imagens dos brancos.

Os valores de precisão sob condições de repetitividade e precisão intermediária, expressos como desvio padrão relativo (DPR_r e DPR_R , respectivamente), foram estimadas por análise de variância dos valores preditos para 18 replicatas de amostras em quatro níveis de concentração distintos.⁴⁷ Os valores de DPR_R variaram entre 5,5 e 8,0%, e os valores de DPR_r variaram entre 4,6 e 5,8%. Todos os valores estimados estão abaixo do limite máximo, calculado através da equação de Horwitz.⁴⁰

Tabela 11 - Parâmetros para avaliação das FOM no modelo proposto para a determinação de Vermelho 40 em balas duras

Figuras de Mérito	Parametro	Valores
Veracidade (n=62)	Erro relativo médio	-0,1%
	DPR repetitividade ^a	5,3 % (6,1 %)
		5,8 % (6,1 %)
		5,1 % (6,3 %)
		4,6 % (6,0 %)
Precisão (n=18)	DPR precisão intermediária ^a	6,2 % (9,1 %)
		5,8 % (9,1 %)
		8,0 % (9,4 %)
		5,5 % (9,0 %)
	Linearidade	Durbin-Watson
Inclinação ^b		0,94±0,02
Intercepto ^b		0,33±0,12
Coeficiente de correlação ^b		0,9669
Faixa de trabalho		22,9 – 78,8 mg.Kg ⁻¹
Seletividade do modelo		----
Sensibilidade do modelo		2,64
Sensibilidade analítica (γ)		4,25
γ^1		0,2
Viés		-0,115±0,6578mg.Kg ⁻¹

^a Resultados para seis amostras em quatro diferentes concentrações

^b Parâmetros do ajuste para a curva de valores de referência vs. valores preditos para as amostras de calibração

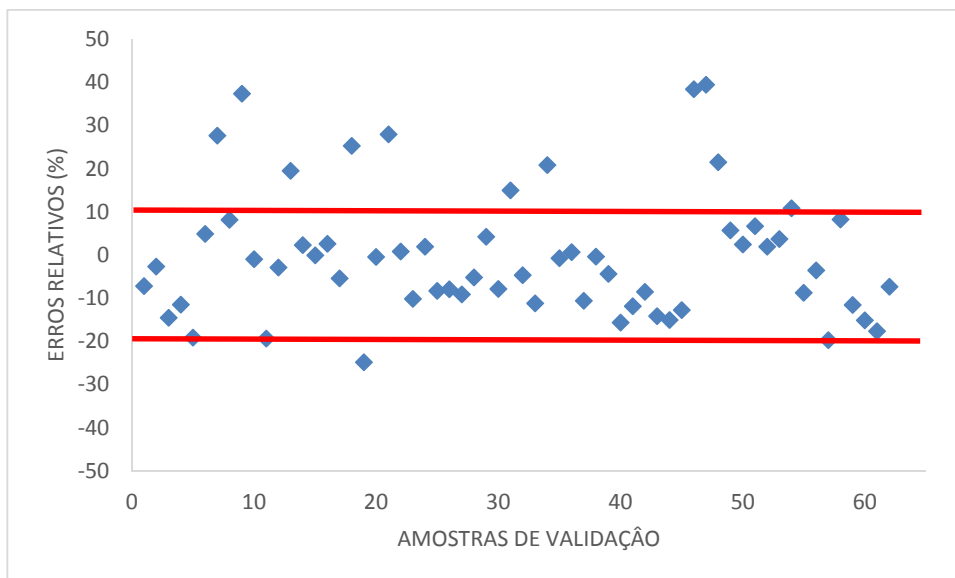


Figura 49 – Erros relativos das amostras do conjunto de validação (Limites estabelecidos pela legislação em vermelho)

Os resultados de veracidade e precisão confirmam a exatidão do método. Considerando a exatidão e a linearidade, a faixa de trabalho do método foi definida como sendo de 22,9 a 78,8 mg Kg⁻¹ de Vermelho 40.

O viés apresentou um valor de $t(1,13)$ menor que o valor crítico (1,67 para 57 graus de liberdade e 95% de confiança), o que indica ausência de erro sistemático nas predições do modelo. O modelo também apresentou valores de RPD dentro do recomendado pela literatura.^{56,57}

6.4.4 β -TI

O β -TI estimado confirmou que existe uma tendência do modelo desenvolvido em fornecer resultados com erros positivos. Como é possível verificar na Figura 52, existe uma maior quantidade de erros positivos, e estes deslocaram o β -TI para fora do limite recomendado pelo guia de validação brasileiro,³⁷ sendo os maiores desvios encontrados nas amostras de mais baixa concentração. Tal condição talvez possa ser explicada pela variação da tonalidade do corante em função da concentração. A cor, diferentemente da absorbância, não varia de maneira linear à medida que aumentamos a concentração de corante, o que pode ser um complicador quando se desenvolve métodos utilizando imagens que possuem uma ampla faixa analítica de trabalho.

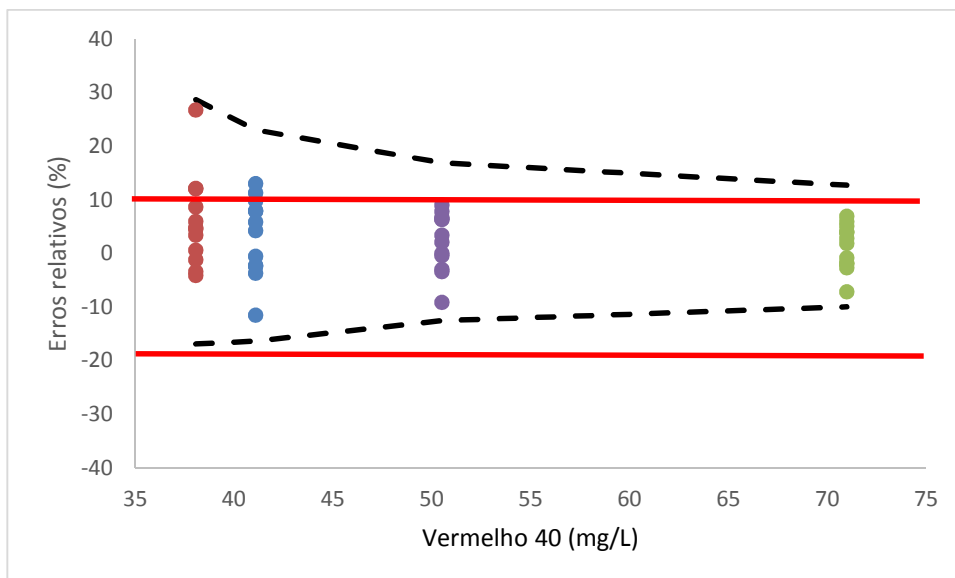


Figura 50 - β -TI estimados para o modelo (linhas tracejadas) e erros máximos relativos definidos pelo guia de validação brasileiro (linhas sólidas)

6.5 Conclusões

Um método de calibração baseado em imagens obtidas através de um telefone celular foi desenvolvido e validado para determinação do corante vermelho 40 em balas duras. Este método foi capaz de quantificar o corante de maneira não destrutiva, além de ser muito mais rápido que o método de referência. O método foi submetido a um processo completo de validação analítica, não atendendo porém, a todos os parâmetros exigidos.

Mesmo não atendendo a todos parâmetros exigidos, a utilização de um simples telefone celular como instrumento analítico capaz de realizar medições da ordem de partes por milhão é um fato relevante. A utilização de imagens digitais em química analítica é um campo relativamente recente, e muitos avanços ainda são necessários, mas o resultados obtidos nesta aplicação provam que este campo é muito promissor.

7 Conclusões Gerais

A aplicação de técnicas espectroscópicas e análise de imagens digitais se mostraram poderosas ferramentas na análise de alimentos. A associações entre estas técnicas e ferramentas quimiométricas como o PLS, PLS-DA e o N-PLS permitiu o desenvolvimento de modelos capaz de determinar parâmetros de qualidade, concentração de aditivos e detectar fraudes de maneira simples, rápida e eficiente.

Um importante entrave que impede métodos multivariados de serem aceitos como métodos oficiais de análise ou mesmo de rotina é a falta de guias específicos que contemplem as particularidades destes métodos. Esta tese é um passo para essa harmonização entre a realidade univariada e a multivariada, tentando propor adaptações para que estes métodos possam ser validados e aplicados de maneira adequada.

8 Referências Bibliográficas

1. Burns, D. A. & Ciurczak, E. W. *Handbook of Near Infrared Analysis*. 3^a. ed. 2008, Boca Raton: CRC Press. 836p.
2. Sun, D. W. *Infrared Spectroscopy for Food Quality Analysis and Control*. 1^a. ed. 2009, Nova York: Academic Press. 451p.
3. Hart, J. R., Norris, K. H. & Golumbic, G. Determination of the moisture content of seeds by near-infrared spectrophotometry of their methanol extracts. *Cereal Chem.* **39**, 94–99 (1962).
4. Pasquini, C. Review near infrared spectroscopy : Fundamentals , practical aspects and analytical applications. *J. Braz. Chem. Soc.* **14**, 198–219 (2003).
5. Pavia, D., Lampman, G., Kriz, G. & Vyvyan, J. *Introduction to Spectroscopy*. 4^a. ed. 2008. Stamford: Cengage Learning. 752p.
6. Karoui, R., Downey, G. & Blecker, C. Mid-infrared spectroscopy coupled with chemometrics: a tool for the analysis of intact food systems and the exploration of their molecular structure-quality relationships - a review. *Chem. Rev.* **110**, 6144–6168 (2010).
7. Karoui, R., Mouazen, A. M., Dufour, E., Pillonel, L., Schaller, E., Picque, D., De Baerdemaeker, J. & Bossier, J. A comparison and joint use of NIR and MIR spectroscopic methods for the determination of some parameters in European Emmental cheese. *Eur. Food Res. Technol.* **223**, 44–50 (2006).
8. Ledley, R. S. High-Speed Automatic Analysis of Biomedical Pictures. *Science* **146**, 216–223 (1964).

9. Geladi, P., Wold, S. & Esbensen, K. Image analysis and chemical information in images. *Anal. Chim. Acta* **191**, 473–480 (1986).
10. Foca, G., Masino, F., Antonelli, A. & Ulrici, A. Prediction of compositional and sensory characteristics using RGB digital images and multivariate calibration techniques. *Anal. Chim. Acta* **706**, 238–245 (2011).
11. Iqbal, Z. & Bjorklund, R. B. Assessment of a mobile phone for use as a spectroscopic analytical tool for foods and beverages. *Int. J. Food Sci. Technol.* **46**, 2428–2436 (2011).
12. Santos, P. M., Wentzell, P. D. & Pereira-Filho, E. R. Scanner digital images combined with color parameters: A case study to detect adulterations in liquid cow's milk. *Food Anal. Meth.* **5**, 89–95 (2012).
13. Borin, A., Ferrão, M. F., Mello, C., Cordi, L., Pataca, L. C., Durán, N. & Poppi, R. J. Quantification of *Lactobacillus* in fermented milk by multivariate image analysis with least-squares support-vector machines. *Anal. Bioanal. Chem.* **387**, 1105–1112 (2007).
14. Acevedo, C. A., Skurtys, O., Young, M. E., Enrione, J., Pedreschi, F. & Osorio, F. A non-destructive digital imaging method to predict immobilized yeast-biomass. *LWT - Food Sci. Technol.* **42**, 1444–1449 (2009).
15. Godinho, M. S., Oliveira, A. E. & Sena, M. M. Determination of interfacial tension of insulating oils by using image analysis and multi-way calibration. *Microchem. J.* **96**, 42–45 (2010).
16. Oliveira, L. F., Canevari, N. T., Guerra, M. B. B., Pereira, F. M. V., Schaefer, C. E. G. R. & Pereira-Filho, E. R. Proposition of a simple method for chromium (VI) determination in soils from remote places applying digital images: A case study from Brazilian Antarctic Station. *Microchem. J.* **109**, 165–169 (2013).
17. Huang, J., Wium, H., Qvist, K. B. & Esbensen, K. H. Multi-way methods in image analysis—relationships and applications. *Chemom. Intell. Lab. Syst.* **66**, 141–158 (2003).
18. Ferreira, M. M. C., Antunes, A. M., Melgo, M. S. & Volpe, P. L. O. Quimiometria I: Calibração multivariada, um tutorial. *Quim. Nova* **22**, 724–731 (1999).
19. Barros Neto, B. de, Scarminio, I. S. & Bruns, R. E. 25 anos de quimiometria no Brasil. *Quim. Nova* **29**, 1401–1406 (2006).
20. de Sena, M. M., Poppi, R. J., Frighetto, R. T. & Valarini, P. Avaliação do uso de métodos quimiométricos em análise de solos. *Quim. Nova* **23**, 547–556 (2000).

21. Shao, X., Bian, X., Liu, J., Zhang, M. & Cai, W. Multivariate calibration methods in near infrared spectroscopic analysis. *Anal. Meth.* **2**, 1662–1666 (2010).
22. Bro, R. Multiway calibration. Multilinear PLS. *J. Chemom.* **10**, 47–61 (1996).
23. Bro, R. Multi-way Analysis in the Food Industry: models, algorithms and applications. 309p. Tese de doutorado. Real Universidade de Veterinária e Agricultura, Dinamarca (1998).
24. Barker, M. & Rayens, W. Partial least squares for discrimination. *J. Chemom.* **17**, 166–173 (2003).
25. Rodriguez-Otero, J. L., Hermida, M. & Centeno, J. Analysis of Dairy Products by Near-Infrared Spectroscopy: A Review. *J. Agric. Food Chem.* **45**, 2815–2819 (1997).
26. Sun, T., Huang, K., Xu, H. & Ying, Y. Research advances in nondestructive determination of internal quality in watermelon/melon: A review. *J. Food Eng.* **100**, 569–577 (2010).
27. Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I. & Lammertyn, J. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biol. Technol.* **46**, 99–118 (2007).
28. Bucci, R., Magri, A. D., Marini, D. & Marini, F. Chemical authentication of extra virgin olive oil varieties by supervised chemometric procedures. *J. Agric. Food Chem.* **50**, 413–418 (2002).
29. García-Alvarez, M., Ceresuela, S., Huidobro, J. F., Hermida, M. & Rodríguez-Otero, J. L. Determination of polarimetric parameters of honey by near-infrared transmittance spectroscopy. *J. Agric. Food Chem.* **50**, 419–425 (2002).
30. Llario, R., Iñón, F.A., Garrigues, S. & de la Guardia, M. Determination of quality parameters of beers by the use of attenuated total reflectance-Fourier transform infrared spectroscopy. *Talanta* **69**, 469–480 (2006).
31. Antonelli, A., Cocchi, M., Fava, P., Foca, G., Franchini, G. C., Manzini, D. & Ulrici, A. Automated evaluation of food colour by means of multivariate image analysis coupled to a wavelet-based classification algorithm. *Anal. Chim. Acta* **515**, 3–13 (2004).
32. Haughey, S. A., Graham, S. F., Cancouët, E. & Elliott, C. T. The application of Near-Infrared Reflectance Spectroscopy (NIRS) to detect melamine adulteration of soya bean meal. *Food Chem.* **136**, 1557–1561 (2013).

33. Gurdeniz, G. & Ozen, B. Detection of adulteration of extra-virgin olive oil by chemometric analysis of mid-infrared spectral data. *Food Chem.* **116**, 519–525 (2009).
34. Guerreiro, J. S., Barros, M., Fernandes, P., Pires, P. & Bardsley, R. Principal component analysis of proteolytic profiles as markers of authenticity of PDO cheeses. *Food Chem.* **136**, 1526–1532 (2013).
35. Casale, M., Oliveri, P., Casolino, C., Sinelli, N., Armanino, C., Forina, M. & Lanteri, S. Characterisation of PDO olive oil Chianti Classico by non-selective (UV-visible, NIR and MIR spectroscopy) and selective (fatty acid composition) analytical techniques. *Anal. Chim. Acta* **712**, 56–63 (2012).
36. Alves, M. R., Cunha, S. C., Amaral, J. S., Pereira, J. A. & Oliveira, M. B. Classification of PDO olive oils on the basis of their sterol composition by multivariate analysis. *Anal. Chim. Acta* **549**, 166–178 (2005).
37. Secretaria de Defesa Agropecuária. Ministério da Agricultura, Pecuária e Abastecimento. *Instrução Normativa n°24 - Anexo II.* (2009).
38. Secretaria de Defesa Agropecuária. Ministério da Agricultura, Pecuária e Abastecimento. Manual de garantia da qualidade analítica: Resíduos e Contaminantes em alimentos. (2011).
39. Instituto Nacional de Metrologia, Normalização e Qualidade Industrial - INMETRO. - “Orientações sobre validação de métodos de ensaios químicos”. DOQ-CGCRE-008, Rio de Janeiro, 2011. Revisão 04, 20p.
40. Food and Agriculture Organization. Validation of analytical methods for food control. Report of a Joint FAO/IAEA Expert Consultation, FAO Food and Nutrition Paper n° 68, Roma, Itália.
41. Thompson, M., Ellison, S. L. R. & Wood, R. Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC Technical Report). *Pure Appl. Chem.* **74**, 835–855 (2002).
42. AOAC (Association of Official Analytical Chemists). AOAC Official Methods of analysis. Guidelines for Standard Method Performance Requirements. Washington: AOAC, 2012.
43. EURACHEM. The fitness for purpose of analytical methods, a laboratory guide to method validation and related topics. Teddington: LGC, 1998, 61p.
44. Bro, R. & Andersen, C. M. Theory of net analyte signal vectors in inverse regression. *J. Chemom.* **17**, 646–652 (2003).
45. 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. *J. Agric. Food Chem.* **55**, 8331–8338 (2007).
46. Annual Book of ASTM Standards, Standard Practices for Infrared Multivariate Quantitative Analysis - E1655-12, ASTM International, West Conshohocken, Pennsylvania, EUA, 2012.
 47. Silva, M. A. M., Ferreira, M. H., Braga, J. W. B. & Sena, M. M. Development and analytical validation of a multivariate calibration method for determination of amoxicillin in suspension formulations by near infrared spectroscopy. *Talanta* **89**, 342–351 (2012).
 48. De Souza, S. V. C. & Junqueira, R. G. A procedure to assess linearity by ordinary least squares method. *Anal. Chim. Acta* **552**, 25–35 (2005).
 49. Lorber, A. Error propagation and figures of merit for quantification by solving matrix equations. *Anal. Chem.* **58**, 1167–1172 (1986).
 50. Lorber, A., Faber, K. & Kowalski, B. R. Net analyte signal calculation in multivariate calibration. *Anal. Chem.* **69**, 1620–1626 (1997).
 51. Olivieri, A. C., Faber, N. M., Ferré, J., Boqué, R., Kalivas, J. H. & Mark, H. Uncertainty estimation and figures of merit for multivariate calibration (IUPAC Technical Report). *Pure Appl. Chem.* **78**, 633–661 (2006).
 52. Faber, N. M. Response to “Comments on construction of confidence intervals in connection with partial least squares.” *J. Chemom.* **14**, 363–369 (2000).
 53. Faber, N. M. Exact presentation of multivariate calibration model as univariate calibration graph. *Chemom. Intell. Lab. Syst.* **50**, 107–114 (2000).
 54. Instituto Nacional De Metrologia, Qualidade E Tecnologia; Oliveira, J.C.V. (coord.). Vocabulário Internacional de Metrologia: conceitos fundamentais e gerais e termos associados. Rio de Janeiro: INMETRO, 2009.
 55. Durbin, J. & Watson, G. S. Testing For Serial Correlation in Least Squares Regression. *Biometrika* **38**, 159–178 (1951).
 56. Valderrama, P. & Braga, J. W. B. Estado da arte de figuras de mérito em calibração multivariada. *Quim. Nova*, **32**, 1278–1287 (2009).
 57. Williams, P & Norris, K. Near-infrared technology in the agricultural and food industries, 1^ª ed. 2001. American Association of Cereal Chemists Inc.: St. Paul.
 58. Patil, A. G., Oak, M. D., Taware, S. P., Tamhankar, S. A. & Rao, V. S. Nondestructive estimation of fatty acid composition in soybean [Glycine

- max (L.) Merrill] seeds using Near-Infrared Transmittance Spectroscopy. *Food Chem.* **120**, 1210–1217 (2010).
59. Faber, N. M. & Bro, R. Standard error of prediction for multiway PLS. *Chemom. Intell. Lab. Syst.* **61**, 133–149 (2002).
 60. de Vries, S. & J.F. ter Braak, C. Prediction error in partial least squares regression: a critique on the deviation used in The Unscrambler. *Chemom. Intell. Lab. Syst.* **30**, 239–245 (1995).
 61. Faber, K. & Kowalski, B. R. Prediction error in least squares regression: Further critique on the deviation used in The Unscrambler. *Chemom. Intell. Lab. Syst.* **34**, 283–292 (1996).
 62. Zhang, L. & Garcia-Munoz, S. A comparison of different methods to estimate prediction uncertainty using Partial Least Squares (PLS): A practitioner's perspective. *Chemom. Intell. Lab. Syst.* **97**, 152–158 (2009).
 63. Rozet, E., Wascotte, V., Lecouturier, N., Pr at, V., Dew e, W., Boulanger, B. & Hubert, P. Improvement of the decision efficiency of the accuracy profile by means of a desirability function for analytical methods validation. Application to a diacetyl-monoxime colorimetric assay used for the determination of urea in transdermal iontophoretic ext. *Anal. Chim. Acta* **591**, 239–247 (2007).
 64. Mantanus, J., Zi mons, E., Rozet, E., Streel, B., Klinkenberg, R., Evrard, B., Rantanen, J. & Hubert, P. Building the quality into pellet manufacturing environment--feasibility study and validation of an in-line quantitative near infrared (NIR) method. *Talanta* **83**, 305–311 (2010).
 65. Valc rcel, M. & C rdenas, S. Modern qualitative analysis. *TrAC Trends Anal. Chem.* **24**, 467 (2005).
 66. Magalh es, C. G., De Paiva, C. R., Botelho, B. G., De Oliveira, A. M., De Souza, L. F., Nonaka, C. V., Santos, K. V., Farias, L. M. & Carvalho, M. A. In-house validation of Premi@Test, a microbiological screening test with solvent extraction, for the detection of antimicrobial residues in poultry muscles. *Food Addit. Contam. Part A* **29**, 535–540 (2011).
 67. Pikkemaat, M. G. Microbial screening methods for detection of antibiotic residues in slaughter animals. *Anal. Bioanal. Chem.* **395**, 893–905 (2009).
 68. European Community Directive 657. Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results, Official Journal, L221, 8-36, Bruxelles, B lgica.
 69. Trullols, E., Ruis nchez, I. & Rius, F. X. Validation of qualitative analytical methods. *TrAC Trends Anal. Chem.* **23**, 137–145 (2004).

70. Gondim, C. S., Junqueira, R. G. & Souza, S. V. C. Tendências em validação de métodos de ensaios qualitativos (Trends in implementing the validation of qualitative methods of analysis). *Rev. Inst. Adolfo Lutz* **70**, 433–447 (2011).
71. NATA. Technical Note 17. Guidelines for the validation and verification of quantitative and qualitative test methods. 32p. Canberra: National Association of Testing Authorities (2013).
72. Gondim, C. S., Coelho, O. A., Alvarenga, R. L., Junqueira, R. G. & de Souza, S. V. An appropriate and systematized procedure for validating qualitative methods: its application in the detection of sulfonamide residues in raw milk. *Anal. Chim. Acta* **9**, 11–22 (2014).
73. Botelho, B., Mendes, B. P. & Sena, M. Development and Analytical Validation of Robust Near-Infrared Multivariate Calibration Models for the Quality Inspection Control of Mozzarella Cheese. *Food Anal. Meth.* **6**, 881–891 (2013).
74. Pulido, A., Ruisánchez, I., Boqué, R. & Rius, F. X. Uncertainty of results in routine qualitative analysis. *TrAC Trends Anal. Chem.* **22**, 647–654 (2003).
75. Wise, B. W., Gallagher, N. B., Bro, R., Shaver, J. M., Windig, W. & Koch, R. S. PLS-Toolbox 4.0 for use with Matlab™ (Manual). (2006).
76. Langton, S. D., Chevenement, R., Nagelkerke, N. & Lombard, B. Analysing collaborative trials for qualitative microbiological methods: accordance and concordance. *Int. J. Food Microbiol.* **79**, 175–181 (2002).
77. Harbutt, J., *The World Cheese*. 1^a. ed. 2009, Nova Iorque: DK Publishing
78. Teixeira, L. V & Fonseca, L. M. Perfil físico-químico do soro de queijos mozzarella e minas-padrão produzidos em várias regiões do estado de Minas Gerais. *Arq. Bras. Med. Vet. Zootec.* **60**, 243-250 (2008).
79. Secretaria de Defesa Agropecuária. Ministério da Agricultura, Pecuária e Abastecimento. Portaria n° 364, Brasil (1997)
80. Y. Pomeranz & C. E. Meloan: *Food Analysis: Theory and Practice*. 3^a. ed. 2002, Berlim: Springer Science & Business Media
81. Badertscher, R., Berger, T. & Kuhn, R. Densitometric determination of the fat content of milk and milk products. *Int. Dairy J.* **17**, 20–23 (2007).
82. Bradley, R. L. J. & Vanderwarn, M. A. Determination of moisture in cheese and cheese products. *J. AOAC Int.* **84**, 570–592 (2001).
83. Gera, L. B. & Norris, K. H. Influence of fat concentration on the absorption spectrum of milk in the near infrared region. *Isr. J. Agric. Res.* **18**, 117-124 (1968).

84. Pierce, M. M. & Wehling, R. L. Comparison of sample handling and data treatment methods for determining moisture and fat in Cheddar cheese by near-infrared spectroscopy. *J. Agric. Food Chem.* **42**, 2830–2835 (1994).
85. Rodriguez-Otero, J. L., Hermida, M. & Cepeda, A. Determination of fat, protein, and total solids in cheese by near-infrared reflectance spectroscopy. *J. AOAC Int.* **78**, 802–806 (1995).
86. Wittrup, C. & Nørgaard, L. Rapid Near Infrared Spectroscopic Screening of Chemical Parameters in Semi-hard Cheese Using Chemometrics. *J. Dairy Sci.* **81**, 1803–1809 (1998).
87. Blazquez, C., Downey, G., Donnell, C. O. & Callaghan, O. Prediction of moisture, fat and inorganic salts in processed cheese by near infrared reflectance spectroscopy and multivariate data analysis. **157**, 149–157 (2004).
88. Lucas, A., Andueza, D., Rock, E. & Martin, B. Prediction of Dry matter, fat, ph, vitamins, minerals, carotenoids, total antioxidant capacity, and color in fresh and freeze-dried cheeses by visible-near-infrared reflectance spectroscopy. *J. Agric. Food Chem.* **56**, 6801–6808 (2008).
89. Pi, F., Shinzawa, H., Ozaki, Y. & Han, D. Non-destructive determination of components in processed cheese slice wrapped with a polyethylene film using near-infrared spectroscopy and chemometrics. *Int. Dairy J.* **19**, 624–629 (2009).
90. Ntsame Affane, A. L., Fox, G. P., Sigge, G. O., Manley, M. & Britz, T. J. Simultaneous prediction of acidity parameters (pH and titratable acidity) in Kefir using near infrared reflectance spectroscopy. *Int. Dairy J.* **21**, 896–900 (2011).
91. Huang, H., Yu, H., Xu, H. & Ying, Y. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *J. Food Eng.* **87**, 303–313 (2008).
92. Ferreira, M. H., Braga, J. W. B. & Sena, M. M. Development and validation of a chemometric method for direct determination of hydrochlorothiazide in pharmaceutical samples by diffuse reflectance near infrared spectroscopy. *Microchem. J.* **109**, 158–164 (2013).
93. Kennard, R. W. & Stone, L. A. Computer Aided Design Of Experiments. *Techometrics* **11**, 137–148 (1969).
94. Isaksson, T. & Naes, T. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. *Appl. Spectrosc.* **42**, 1273–1284 (1988).
95. Workman, J.J. and L. Weyer, Practical guide to interpretative near infrared spectroscopy. 2008, Boca Raton: CRC - Press.

96. Draper, N.R. and H. Smith, *Applied Regression Analysis*. 1998, Nova Jersey: John Wiley & Sons.
97. Horwitz, W. & Albert, R. The Horwitz ratio (HorRat): A useful index of method performance with respect to precision. *J. AOAC Int.* **89**, 1095–1109 (2006).
98. Resende, M.F.S., *Queijo Minas Artesanal da Serra da Canastra: Influência da altitude e do nível de cadastramento das queijarias nas características físico-químicas e microbiológicas*. 72p. Dissertação de mestrado. Universidade Federal de Minas Gerais, Brasil. (2010).
99. Nogueira, M. C. L., Lubachevsky, G. & Rankin, S. A. A study of the volatile composition of Minas cheese. *LWT - Food Sci. Technol.* **38**, 555–563 (2005).
100. MINAS GERAIS. Assembléia Legislativa do Estado de Minas Gerais. Decreto nº 44.864 de 01 de agosto de 2008. Altera o regulamento da lei nº 14.185 de 31 de janeiro de 2002, que dispõe sobre o processo de produção do queijo Minas artesanal. Belo Horizonte: Assembléia Legislativa.
101. MINAS GERAIS. Assembléia Legislativa do Estado de Minas Gerais. Decreto nº 42.645, de 05 de junho de 2002. Aprova o regulamento da Lei nº 14.185, de 31/01/2002, que dispõe sobre o processo de produção de queijo Minas artesanal.
102. Fox, P. F. & McSweeney, P. L. H. *Advanced Dairy Chemistry: Volume 1: Proteins, Parts A&B: Protein*. 1ª ed. 2003, Springer Science & Business Media:Berlim.1236p
103. EMBRAPA - Empresa Brasileira de Pesquisa Agrícola. Estatísticas de produção de leite. (2014). Disponível em <<http://www.cnpqgl.embrapa.br/nova/informacoes/estatisticas/producao/producao.php>>. Acesso em julho/2014
104. Hoorfar, J. *Case Studies in Food Safety and Authenticity: Lessons From Real-Life Situations*. 1ª ed. 2012, Woodhead Publishing Limited:Cambridge (2012)
105. UOL, Operação desmonta esquema de empresas que adulteravam leite no Sul. (2014). Disponível em <<http://noticias.uol.com.br/cotidiano/ultimas-noticias/2013/05/08/operacao-desmonta-esquema-de-empresas-que-fraudavam-leite-no-sul.htm>>
106. Souza, S. S., Cruz, A. G., Walter, E. H. M., Faria, J. A. F., Celeghini, R. M. S., Ferreira, M. M. C., Granato, D. & Sant'Ana, A. S. Monitoring the authenticity of Brazilian UHT milk : A chemometric approach. *Food. Chem.* **124**, 692-695 (2011)

107. IAL. Métodos Físicos-Químicos para Análise de Alimentos. 4^a ed. Instituto Adolfo Lutz: São Paulo (IAL), (2005).1020p
108. MAPA. Instrução Normativa número 22 de 19 de abril de 2006. (2006).
109. Kasemsumran, S., Thanapase, W. & Kiatsoonthon, A. Feasibility of near-infrared spectroscopy to detect and to quantify adulterants in cow milk. *Anal. Sci.* **23**, 907–910 (2007).
110. Nicolaou, N., Xu, Y. & Goodacre, R. Fourier transform infrared spectroscopy and multivariate analysis for the detection and quantification of different milk species. *J. Dairy Sci.* **93**, 5651–5660 (2014).
111. Santos, P. M., Pereira-Filho, E. R. & Rodriguez-Saona, L. E. Application of hand-held and portable infrared spectrometers in bovine milk analysis. *J. Agric. Food Chem.* **61**, 1205–1211 (2013).
112. Santos, P. M., Pereira-Filho, E. R. & Rodriguez-Saona, L. E. Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis. *Food Chem.* **138**, 19–24 (2013).
113. Zhang, L. G., Zhang, X., Ni, L., Xue, Z., Gu, X. & Huang, S. Rapid identification of adulterated cow milk by non-linear pattern recognition methods based on near infrared spectroscopy. *Food Chem.* **145**, 342–348 (2014).
114. Kizil, R., Irudayaraj, J. & Seetharaman, K. Characterization of Irradiated Starches by Using FT-Raman and FTIR Spectroscopy. *J. Agric. Food Chem.* **50**, 3912–3918 (2002).
115. Tewari, J. C. & Malik, K. In situ laboratory analysis of sucrose in sugarcane bagasse using attenuated total reflectance spectroscopy and chemometrics. *Int. J. Food Sci. Technol.* **42**, 200–207 (2007).
116. Brown, C. D. & Green, R. L. Critical factors limiting the interpretation of regression vectors in multivariate calibration. *TrAC Trends Anal. Chem.* **28**, 506–514 (2009).
117. Chong, I.-G. & Jun, C.-H. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* **78**, 103–112 (2005).
118. Ghoreishi, S. M., Behpour, M. & Golestaneh, M. Simultaneous determination of Sunset yellow and Tartrazine in soft drinks using gold nanoparticles carbon paste electrode. *Food Chem.* **132**, 637–641 (2012).
119. Xing, Y., Xing, Y., Meng, M., Xue, H., Zhang, T., Yin, Y & Xi, R. Development of a polyclonal antibody-based enzyme-linked

- immunosorbent assay (ELISA) for detection of Sunset Yellow FCF in food samples. *Talanta* **99**, 125–131 (2012).
120. ANVISA. Agência Nacional de Vigilância Sanitária. Compêndio da Legislação Brasileira de Aditivos Alimentares, Brasília. 163p. (2011)
 121. Bonan, S., Fedrizzi, G., Menotta, S. & Elisabetta, C. Simultaneous determination of synthetic dyes in foodstuffs and beverages by high-performance liquid chromatography coupled with diode-array detector. *Dyes Pigm.* **99**, 36–40 (2013).
 122. Vidotti, E. C., Costa, W. F. & Oliveira, C. C. Development of a green chromatographic method for determination of colorants in food samples. *Talanta* **68**, 516–521 (2006).
 123. Nevado, J. J. B., Flores, J. R. & Llerena, M. J. V. Square wave adsorptive voltammetric determination of sunset yellow. *Talanta* **44**, 467–474 (1997).
 124. El-Shahawi, M. S., Hamza, A., Al-Sibaai, A. A., Bashammakh, A. S. & Al-Saidi, H. M. A new method for analysis of sunset yellow in food samples based on cloud point extraction prior to spectrophotometric determination. *J. Ind. Eng. Chem.* **19**, 529–535 (2013).
 125. Berzas Nevado, J. J., Rodríguez Flores, J., Guiberteau Cabanillas, C., Villaseñor Llerena, M. J. & Contento Salcedo, a. Resolution of ternary mixtures of Tartrazine, Sunset yellow and Ponceau 4R by derivative spectrophotometric ratio spectrum-zero crossing method in commercial foods. *Talanta* **46**, 933–942 (1998).
 126. Nevado, J., Flores, J. R., Llerena, M. J. V & Fariñas, N. R. Simultaneous spectrophotometric determination of tartrazine, patent blue V, and indigo carmine in commercial products by partial least squares and principal component regression methods. *Talanta* **48**, 895–903 (1999).
 127. El-Sheikh, A. H. & Al-Degs, Y. S. Spectrophotometric determination of food dyes in soft drinks by second order multivariate calibration of the absorbance spectra-pH data matrices. *Dyes Pigm.* **97**, 330–339 (2013).
 128. Pedjie, N. Analysis of color additives in beverages with the PerkinElmer Flexar FX-15 system equipped with a PDA detector. Application Note, PerkinElmer, Waltham, USA. Disponível em <http://www.perkinelmer.com/CMSResources/Images/44-134896APP_Food_Additives_in_Beverages.pdf>
 129. Benassi, M. D. T. Nota Prévia – Metodologia para avaliação de cristalização em balas duras por imagem digitalizada. *Cienc. Tecnol. Aliment.* **9**, 151–155 (2006)
 130. ANVISA. Agência Nacional de Vigilância Sanitária. Resolução nº44, de 05 de Agosto de 1977.

131. Pinheiro, S., Schumann, A., Lucia, M. & Polônio, T. Avaliação do consumo de corantes artificiais por lactentes , pré-escolares e escolares. *Cienc. Tecnol. Aliment.* **28**, 534–539 (2008).
132. Pinheiro, M. C. O. & Abrantes, S. M. P. Avaliação da exposição aos corantes artificiais em balas e chicletes por crianças entre 3 e 9 anos estudantes de escolas particulares da Tijuca/Rio de Janeiro. *Rev. Anal. Disponível em <<http://www.revistaanalytica.com.br/artigos/8.html>>*.
133. Yilmaz, U. T., Ergun, F. & Yilmaz, H. Determination of the food dye carmine in milk and candy products by differential pulse polarography. *J. Food Drug Anal.* **22**, 329–335 (2014).
134. Prado, M. A. & Godoy, H. T. Teores de corantes artificiais em alimentos determinados por cromatografia líquida de alta eficiência . *Quim. Nova* **30** , 268–273 (2007).
135. Asadpour-Zeynali, K., Maryam Sajjadi, S., Taherzadeh, F. & Rahmanian, R. Analysis of variation matrix array by bilinear least squares-residual bilinearization (BLLS-RBL) for resolving and quantifying of foodstuff dyes in a candy sample. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **123**, 273–281 (2014).
136. Preechaburana, P., Suska, A. & Filippini, D. Biosensing with cell phones. *Trends Biotechnol.* **32**, 351–355 (2014).
137. Intaravanne, Y., Sumriddetchkajorn, S. & Nukeaw, J. Cell phone-based two-dimensional spectral analysis for banana ripeness estimation. *Sensors Actuators B Chem.* **168**, 390–394 (2012).
138. Hazir, M. H. M., Shariff, A. R. M. & Amiruddin, M. D. Determination of oil palm fresh fruit bunch ripeness—Based on flavonoids and anthocyanin content. *Ind. Crops Prod.* **36**, 466–475 (2012).
139. Kanwar, L. & Rao, K. K. Development of a low-cost portable colorimeter for the estimation of fluoride in drinking water. *Sensors Actuators B Chem.* **149**, 245–251 (2010).
140. García, A., Erenas, M.M., Marinetto, E. D., Abad, C. A., de Orbe-Paya, I., Palma, A. J. & Capitán-Vallvey, L. F. Mobile phone platform as portable chemical analyzer. *Sensors Actuators B Chem.* **156**, 350–359 (2011).
141. Iqbal, Z. & Bjorklund, R. B. Colorimetric analysis of water and sand samples performed on a mobile phone. *Talanta* **84**, 1118–1123 (2011).
142. Borzelleca, J. F., Olson, J. W. & Reno, F. E. Lifetime toxicity/carcinogenicity studies of FD & C Red No. 40 (allura red) in mice. *Food Chem. Toxicol.* **29**, 313–319 (1991).

143. Messick, N. J., Kalivas, J. H. & Lang, P. M. Selectivity and related measures for nth-order data. *Anal. Chem.* **68**, 1572–1579 (1996).