



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO



DEPARTAMENTO DE CIÊNCIAS DOMÉSTICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA DE
ALIMENTOS

**CLASSIFICAÇÃO MULTIVARIADA DE LEITE UHT QUANTO À PRESENÇA
DE LACTOSE UTILIZANDO ESPECTROSCOPIA NO INFRAVERMELHO
PRÓXIMO (NIR)**

GLAUCIA FERREIRA DE LIMA

Recife

2017

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE CIÊNCIAS DOMÉSTICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA DE
ALIMENTOS

GLAUCIA FERREIRA DE LIMA

**CLASSIFICAÇÃO MULTIVARIADA DE LEITE UHT QUANTO À PRESENÇA
DE LACTOSE UTILIZANDO ESPECTROSCOPIA NO INFRAVERMELHO
PRÓXIMO (NIR)**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos, da Universidade Federal Rural de Pernambuco, como requisito para obtenção do Grau de Mestre em Ciência e Tecnologia de Alimentos.

ORIENTADORA: Prof^ª. Dr^ª. Samara Alvachian Cardoso Andrade

CO-ORIENTADORA: Prof^ª. Dr^ª. Fernanda Araújo Honorato

Recife

2017

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas da UFRPE
Biblioteca Central, Recife-PE, Brasil

L732c Lima, Gláucia Ferreira de.
Classificação multivariada de leite UHT quanto à presença de lactose utilizando espectroscopia no infravermelho próximo / Gláucia Ferreira de Lima. – Recife, 2017.
80 f.: il.

Orientador(a): Samara Alvachian Cardoso Andrade.

Coorientador(a): Fernanda Araújo Honorato.

Dissertação (Programa de Pós-graduação em Ciência e Tecnologia de Alimentos) – Universidade Federal Rural de Pernambuco, Departamento de Ciências Domésticas, Recife, 2017.

Inclui referências.

1. Lactose. 2. NIRS. 3. NIR portátil. 4. Classificação multivariada. I. Andrade, Samara Alvachin Cardoso, orient. II. Honorato, Fernanda Araújo, coorient. III. Título

CDD 664

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE CIÊNCIAS DOMÉSTICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA DE
ALIMENTOS

**CLASSIFICAÇÃO MULTIVARIADA DE LEITE UHT QUANTO À PRESENÇA
DE LACTOSE UTILIZANDO ESPECTROSCOPIA NO INFRAVERMELHO
PRÓXIMO (NIR)**

Por Gláucia Ferreira de Lima

Esta dissertação foi julgada para obtenção do título de Mestre em Ciência e Tecnologia de Alimentos e aprovada em __/__/__ pelo Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos em sua forma final.

Banca Examinadora:

Prof^ª Dr^ª. Celiane Gomes Maia da Silva
Universidade Federal Rural de Pernambuco (UFRPE)

Prof^ª Dr^ª. Claudete Fernandes Pereira
Universidade Federal de Pernambuco (UFPE)

Prof^ª Dr^ª. Erilane de Castro Lima Machado
Universidade Federal Rural de Pernambuco (UFRPE)

AGRADECIMENTOS

Agradeço às professoras Dr^a. Fernanda Araújo Honorato e Dr^a. Samara Alvachian Cardoso Andrade, pela orientação, por todos os ensinamentos e confiança depositados em mim durante a realização deste trabalho;

A todos os membros do Programa de Pós-graduação em Ciência e Tecnologia de Alimentos (PGCTA) da Universidade Federal Rural de Pernambuco e todos aqueles que contribuíram para a realização deste trabalho, meu sincero agradecimento;

Em especial à minha família pelo apoio, por cada aprendizado e por me manter sempre firme em meus propósitos.

RESUMO

O leite com lactose hidrolisada, produzido pela quebra desse carboidrato por enzimas β -galactosidases, foi desenvolvido a partir dos anos 1970 principalmente para atender aos indivíduos com intolerância à lactose. Entretanto, os métodos analíticos mais utilizados para a análise de açúcares não são eficientes para a avaliação qualitativa da lactose em produtos contendo diferentes açúcares, como no leite sem lactose. Por isso, a utilização de métodos mais precisos na identificação de lactose torna-se essencial para a indústria de alimentos. A espectroscopia no infravermelho próximo (NIRS) é uma técnica analítica emergente na indústria de processamento de alimentos devido a características como não ser destrutiva, de resposta analítica rápida, precisa e de baixo custo. Contudo, o uso de análise multivariada de dados é necessário para extrair a informação codificada no espectro. Assim, o objetivo desse trabalho foi desenvolver e avaliar uma metodologia baseada em espectros NIR, obtidos em equipamentos portátil e de bancada, associada a análise multivariada de dados para classificação de leite longa vida quanto à presença de lactose. Para isso, 71 amostras de leite UHT foram adquiridas em comércio local, sendo 30 com lactose e 41 sem lactose. No presente trabalho, os espectros no modo de reflectância difusa foram obtidos em espectrômetros de bancada (FT-IR Spectrum Frontier da marca Perkin Elmer) de alta resolução e portátil (MicroNIR 1700 da marca JDSU Corporation) de baixa resolução. Os modelos de classificação utilizando as técnicas supervisionadas PLS-DA e LDA com seleção de variáveis pelos algoritmos genético (GA) e de projeções sucessivas (SPA) foram desenvolvidos a partir dos espectros pré-processados. Para ambos os equipamentos, bancada e portátil, todos os modelos apresentaram sensibilidade e especificidade de 100%, apenas o modelo SPA-LDA para o equipamento portátil obteve sensibilidade de 80%. Dessa forma, espectros NIR, em especial os obtidos com equipamento portátil, em conjunto com a classificação multivariada de dados mostrou-se eficiente para identificação da presença de lactose em leite UHT.

Palavras-chave: Lactose, NIRS, NIR portátil, classificação multivariada.

ABSTRACT

Lactose hydrolyzed milk, produced by the conversion of this carbohydrate by β -galactosidases, was developed since the 1970s mainly to suit individuals with lactose intolerance. However, the analytical methods most used for the analysis of sugars are inefficient for the qualitative evaluation of lactose in products containing different sugars, such as lactose-free milk. Therefore, the use of more sophisticated methods for this analysis becomes essential for the food industry. Near Infrared Spectroscopy (NIRS) represents an emerging analytical technique for the food industry because of characteristics such as non-destructive and rapid analysis. Nevertheless, the use of multivariate data analysis is necessary to extract the encoded information in the spectra. Thus, the aim of this work was to develop and validate a methodology based on NIR spectra obtained in portable and benchtop equipments, associated with the multivariate analysis for the classification of long-life milk regarding the presence of lactose. For this purpose, 71 samples of UHT regular (30) and lactose-free milk (41) were purchased in local commerce. In the present study, the spectra in the diffuse reflectance mode were obtained on high resolution benchtop (FT-IR Spectrum Frontier from Perkin Elmer) and low resolution portable (MicroNIR 1700 from JDSU Corporation) spectrometers. The classification models using the supervised techniques PLS-DA and LDA with variables selection by genetic algorithms (GA) and successive projections (SPA) were developed from the pre-processed spectra. For both equipment, benchtop and portable, all models presented sensitivity and specificity of 100%, only SPA-LDA model for portable equipment obtained sensitivity in 80%. Therefore, NIR spectra, especially those obtained with portable equipment, combined with multivariate classification was effective for the presence identification of lactose in UHT milk.

Keywords: Lactose, NIRS, Portable NIR, multivariate classification.

Lista de Figuras e Quadros

Figura 1. Expressão da hidrólise enzimática da molécula de lactose.

Figura 2. Fluxograma dos processos tecnológicos utilizados para a produção de leite com lactose hidrolisada. No processo “A” a enzima é adicionada na embalagem após o tratamento térmico e a enzima permanece ativa durante a vida de prateleira. No processo “B” a adição da lactase é feita em batelada antes do tratamento térmico e a lactase não é ativa durante a vida de prateleira.

Figura 3. Reação típica da hidrólise enzimática da lactose com concentração inicial de 5% (m/v) por enzima imobilizada. (Δ) lactose, (\diamond) galactose, (\square) glicose, (\circ) galacto-oligossacarídeo.

Figura 4. Tipos de vibrações moleculares.

Figura 5. Tipos de medições para aquisição do espectro de infravermelho.

Figura 6. Esquema das técnicas de monitoramento do processo de produção. Seta vermelha = transferência da amostra; seta tracejada azul = transferência de dados.

Figura 7. Correção da linha de base pela aplicação da 1ª e 2ª derivadas. A 1ª derivada corrige os deslocamentos da linha de base. A 2ª derivada corrige tanto o deslocamento, como a inclinação constante na linha de base.

Figura 8. Esquema dos métodos de reconhecimento de padrão supervisionado e não supervisionado.

Quadro 1. Parâmetros mínimos de qualidade do leite.

Listas de abreviaturas e siglas

AOAC – Association of Official Analytical Chemists (Associação Oficial dos Químicos Analíticos)

CML – Carboximetil-lisina

FT-NIR – Fourier Transform Near Infrared Spectroscopy (Espectroscopia no Infravermelho Próximo com Transformada de Fourier)

GA – Genetic Algorithm (Algoritmo Genético)

GA-LDA – Genetic Algorithm based Linear Discriminant Analysis (Algoritmo Genético Combinado a Análise Discriminante Linear)

GRAS – Geralmente Reconhecidas como Seguros

HMF – Hidroximetilfurfural

HTST – High Temperature Short Time (Alta Temperatura/Curto Tempo)

iPLS - Interval Partial Least-Squares Regression (Quadrados Mínimos Parciais Por Intervalos)

LChG – Issues of Nutrition of the German Society of Food Chemistry

LDA – Linear Discriminant Analysis (Análise Discriminante Linear)

LV – Latent Variables (Variáveis Latentes)

MSC – Multiplicative Scatter Correction (Correção Multiplicativa de Espalhamento)

NIR – Near Infrared (Infravermelho Próximo)

NIRS – Near Infrared Spectroscopy (Espectroscopia no Infravermelho Próximo)

NW – Norris-Williams

PC – Principal Components (Componentes Principais)

PCA – Principal Component Analysis (Análise de Componentes Principais)

PLS – Partial Least Squares (Quadrados Mínimos Parciais)

PLS-DA – Partial Least Squares Discriminant Analysis (Análise Discriminante pelo Método dos Quadrados Mínimos Parciais)

SIMCA – Soft Independent Modeling of Class Analogy

SG – Savitzky-Golay

SNV – Standard Normal Variate (Variação Normal Padrão)

SPA – Successive Projection Algorithm (Algoritmo de Projeção Sucessiva)

SPA-LDA – Successive Projection Algorithm based Linear Discriminant Analysis
(Algoritmo de Projeção Sucessiva Combinado a Análise Discriminante Linear)

SW – Método Stepwise

SW-LDA – Stepwise Linear Discriminant Analysis

UHT – Ultra High Temperature (Ultra Alta Temperatura)

UVE – Uninformative Variable Elimination (Eliminação de Variáveis Não Informativas)

SUMÁRIO

1. INTRODUÇÃO	10
2. PROBLEMA DE PESQUISA.....	12
3. HIPÓTESE	12
4. REFERENCIAL TEÓRICO.....	12
4.1. Leite: composição, qualidade e beneficiamento	12
4.2. Lactose: aspectos digestivos, bioquímicos e tecnológicos	17
4.3. Espectroscopia no infravermelho próximo (NIR).....	24
4.4. Análise multivariada dos dados	30
4.4.1. Preparação dos dados para análise	31
4.4.2. Reconhecimento de padrão.....	35
4.5. Espectroscopia NIR e Análise Multivariada aplicada à análise de produtos lácteos	42
5. REFERÊNCIAS BIBLIOGRÁFICAS	46
6. ARTIGO ORIGINAL	59
Classificação Multivariada de leite UHT utilizando espectrômetros NIR portátil e de bancada.....	59

1. INTRODUÇÃO

O leite é um alimento fluido secretado por fêmeas de todas as espécies de mamíferos, rico em nutrientes como lactose, lipídios, proteínas, aminoácidos, ureia e creatinina (WALSTRA *et al.*, 2005; BUNACIU *et al.*, 2016). Embora seja um alimento nutricionalmente completo, alguns indivíduos podem apresentar intolerâncias a certos componentes, como ao dissacarídeo predominante no leite. Uma alternativa para essa condição é o consumo de leite e derivados com lactose hidrolisada, que são desenvolvidos desde os anos 1970, quando as primeiras enzimas β -galactosidasas se tornaram disponíveis comercialmente. Esse processo de hidrólise é simples e não requer equipamentos específicos na linha de produção (HARJU *et al.*, 2012).

A demanda por produtos sem lactose pelos consumidores intolerantes criou um segmento de mercado para esse gênero alimentício, cujo controle de qualidade tem que ser assegurado. Todavia, os métodos clássicos utilizados para as análises físico-químicas em leite apresentam alguns inconvenientes, como baixa velocidade analítica e utilização de reagentes e equipamentos que elevam os custos operacionais. Dessa forma, torna-se evidente a necessidade em desenvolver métodos rápidos, precisos, de baixo custo e não prejudiciais ao meio ambiente para avaliação dos principais componentes do leite a fim de acompanhar os avanços da indústria moderna (BUNACIU *et al.*, 2016).

As metodologias empregadas em análises físico-químicas de alimentos evoluíram ao longo dos últimos 100 anos, de análises predominantemente da química úmida para análises por técnicas instrumentais. Isso proporcionou o desenvolvimento de alternativas mais precisas, que fornecem dados quantitativos e qualitativos de forma rápida e permitem avaliar diferentes características físico-químicas dos alimentos, reduzindo a subjetividade, a fragilidade e a complexidade dos métodos pioneiros na determinação da composição de alimentos (GALLO & FERRANTI, 2016; MCGORRIN, 2009).

O desenvolvimento de técnicas analíticas mais sofisticadas e acessíveis para a autenticação de produtos alimentares continua em ritmo acelerado, concomitante ao aumento da sensibilização dos consumidores para questões de autenticidade nos alimentos e ao contínuo interesse das autoridades reguladoras e da indústria em garantir que os alimentos estejam corretamente rotulados (REID *et al.*, 2006, KAROUI & BAERDEMAEKER, 2007).

Análises espectroscópicas utilizando radiação infravermelha são úteis para monitorar e avaliar a composição e a qualidade na indústria de alimento desde o início dos anos 60. Nos últimos 40 anos, a espectroscopia no infravermelho próximo (*Near infrared spectroscopy* – NIRS) tem se tornado uma técnica mais atrativa em comparação às técnicas tradicionais devido as suas características não invasivas, permitindo caracterizações quantitativas e qualitativas rápidas e simultâneas, revelando propriedades físicas e químicas de produtos naturais e seus constituintes (HUCK, 2015).

A espectroscopia NIR apresenta-se com popularidade crescente na indústria de processamento de alimentos devido ao seu baixo custo de manutenção e por não necessitar de preparação prévia da amostra, além de ser ambientalmente amigável e não destrutiva (POREP *et al.*, 2015). Entretanto, espectros NIR que fornecem informações estruturais mais complexas relacionadas ao comportamento vibracional das ligações químicas, são de difícil interpretação, pois as bandas de absorção são resultantes de sobretons ou combinações das bandas fundamentais no infravermelho médio (KAROUI & BAERDEMAEKER, 2007; ROGGO *et al.*, 2007). Desta forma, é necessário o uso de técnicas multivariadas de análise para extrair informação relevante contida nos espectros.

A Quimiometria é o conjunto de técnicas de análise multivariada que associa métodos matemáticos e estatísticos otimizados para a seleção de procedimentos experimentais e de tratamento de dados multivariados, permitindo a compreensão adequada da informação química contida na resposta instrumental, como por exemplo, em espectros NIR (ROGGO *et al.*, 2007).

Em vista da contínua demanda de controle de qualidade que exigem determinações rápidas, precisas, de baixo custo e com menor geração de resíduos químicos, o objetivo deste trabalho foi desenvolver e avaliar uma metodologia utilizando espectroscopia NIR, em equipamentos portátil e de bancada, em conjunto à análise multivariada de dados para classificação de leite UHT quanto a presença de lactose.

2. PROBLEMA DE PESQUISA

A classificação de leite UHT em relação à presença de lactose pode ser realizada aplicando-se a espectroscopia no infravermelho próximo utilizando equipamentos de bancada e portátil associada aos métodos de análise multivariada?

3. HIPÓTESE

A partir dos métodos de análise multivariada aplicados aos espectros obtidos por espectroscopia na região do infravermelho próximo com equipamentos de bancada e portátil a classificação de leites UHT em relação à presença de lactose pode ser realizada de forma rápida, precisa e não destrutiva nas amostras.

4. REFERENCIAL TEÓRICO

4.1. Leite: composição, qualidade e beneficiamento

O leite é considerado uma importante fonte de nutrientes, como proteínas, minerais e vitaminas (OCHI *et al.*, 2010). O termo genérico refere-se ao leite de vaca, obtido de animais saudáveis, excluída a secreção láctea entre 15 dias antes e 5 dias após o parto ou até que esteja quase completamente livre de colostro, composto por 87% de água, 4 a 5% de lactose, 3% de proteína, 3 a 4% de gordura, 0,8% de minerais e 0,1% de vitaminas, em média (PEREIRA, 2014).

Nessa mistura biológica complexa de componentes, os lipídios encontram-se emulsionados em glóbulos, as proteínas estão em dispersão coloidal em micelas, a maioria

dos minerais e a lactose apresentam-se em solução, características que fazem com que o leite tenha uma composição singular (HAUG *et al.*, 2007).

A maior parte dos lipídios no leite está presente em glóbulos que variam entre diâmetros menores que 0,5 a maiores do que 8 μm , cada glóbulo revestido por uma membrana lipoproteica, também conhecida por membrana do glóbulo de gordura do leite, é constituído por um núcleo rico em triacilglicerol sintetizado no retículo endoplasmático rugoso da célula epitelial mamária (MATHER, 2011).

As proteínas do leite são uma mistura heterogênea dividida em dois grupos básicos, as caseínas e as proteínas do soro do leite. As caseínas são as proteínas mais abundantes no leite representando 79% das proteínas totais e compreendem as frações α_{s1} -, α_{s2} -, β -, κ - e γ -, diferenciadas pela concentração, conteúdo de fósforo, composição de aminoácidos, peso molecular e ponto isoelétrico. As proteínas do soro do leite são proteínas globulares que constituem aproximadamente 0,6% da composição total do leite e incluem as β -lactoglobulinas, α -lactoalbuminas, lactoferrina, lactoperoxidase, lisozima, soroalbuminas, imunoglobulinas, transferrina e proteose-peptonas, consideradas proteínas solúveis (MICIŃSKI *et al.*, 2013). Ambos os grupos são classificados como proteínas de alto valor biológico contemplando as necessidades em aminoácidos essenciais, a digestibilidade e a biodisponibilidade (PEREIRA, 2014).

Os minerais não apresentam concentrações constantes e ocorrem em equilíbrio químico entre os íons livres e componentes complexos, como proteínas, lipídios, carboidratos ou ligantes de baixo peso molecular, por exemplo, citrato e aminoácidos (SOLA-LARRAÑAGA & NAVARRO-BLASCO, 2009). Esses elementos são classificados em solúveis, quando presentes no soro, e insolúveis, quando associados às micelas de caseína. Minerais como cálcio, fósforo e magnésio exercem papéis importantes na estrutura e estabilidade das micelas proteicas, essenciais para as propriedades tecnológicas do leite (POULSEN *et al.*, 2015).

A lactose (O- β -D-galactopiranosil-(1 \rightarrow 4)- β -D-glucopiranosose, $\text{C}_{12}\text{H}_{22}\text{O}_{11}$) é um dissacarídeo composto por uma molécula de glicose ligada a uma molécula de galactose, sintetizada pelas células epiteliais mamárias a partir de precursores absorvidos pelo sangue, presente em equilíbrio nas formas anoméricas α - e β - em soluções aquosas, com menor solubilidade quando comparada a outros dissacarídeos e doçura próxima a 20% da doçura relativa à sacarose (GÄNZLE *et al.*, 2008; ILLANES, 2016).

Carboidrato predominante no leite, a lactose tem como principal precursor a glicose, dependendo também de outros metabólitos como ATP (adenosina trifosfato) e NADPH (nicotinamida-adenina-dinucleotídeo-fosfato reduzida), para a sua síntese. A produção de leite depende da síntese de lactose mamária devido a sua propriedade osmorregulatória para a absorção de água pelas células secretórias mamárias (LIU *et al.*, 2013). O leite contém também 0,3 a 0,6% de galacto-oligossacarídeos, importantes fontes de energia para o crescimento da variedade de *Lactobacillus bifidus* predominante na microbiota intestinal de lactentes (SHENDURSE & KHEDKAR, 2016).

O leite fresco possui sabor delicado ligeiramente doce, pouco aroma, paladar e sabor residual agradáveis, mostrando-se fortemente influenciado por fatores ambientais, genéticos (THOMAS, 1981), e do manejo animal, tais como raça, estágio de lactação e dieta, além de fatores do processamento térmico (YAYOTA *et al.*, 2013). O leite cru abriga uma microbiota complexa e os atributos de elevado valor nutricional, alto teor de água e pH próximo a neutralidade são condições que favorecem o crescimento de muitos microrganismos, incluindo bactérias de relevância tecnológica, como as bactérias lácteas, e bactérias deteriorantes e patogênicas, aspectos que exercem importante influência nos parâmetros de qualidade e segurança do leite (QUIGLEY *et al.*, 2011).

O conceito de qualidade é complexo e inclui a segurança alimentar, nutrição e atributos relacionados à diferenciação comercial dos produtos, complementando a necessidade de determinação de padrões e regulamentação para que os parâmetros de qualidade sejam alcançados (FARINA *et al.*, 2005). A maioria dos países produtores de laticínios dispõe de regulamentação que orientam a qualidade aceitável do leite (ELLIOTT *et al.*, 1974).

A União Europeia estabelece os padrões de qualidade do leite em 3,5% de gordura, mínimo de 2,9% de proteínas, densidade não inferior a 1,028 g L⁻¹ (a 20 °C), extrato seco desengordurado não inferior a 8,5% e índice crioscópico menor que -0,52 °C (EC, 1992; EC, 1997). A legislação brasileira vigente está em conformidade a esses valores, como pode ser observado no Quadro 1, fixando requisitos mínimos para determinação do padrão de identidade e qualidade do leite (BRASIL, 1996; BRASIL, 2011).

Quadro 1. Parâmetros mínimos de qualidade do leite segundo MAPA.

	Leite para consumo*	Leite cru refrigerado**	Leite UHT***		
			Integral	Semi ou parcialmente desnatado	Desnatado
Teor de gordura (m/m)	3,5%	mín. de 3%	mín. de 3%	0,6 a 2,9	máx. de 0,5
Teor de proteína (m/m)	mín. de 2,9%	mín. de 2,9%			
Acidez (g ácido láctico/100 mL)		0,14 a 0,18	0,14 a 0,18	0,14 a 0,18	0,14 a 0,18
Estabilidade ao etanol 68% (v/v)			estável	estável	estável
Densidade	não inferior a 1,028 g/L a 20 °C	entre 1.028 e 1,034 g/mL a 15/15 °C			
Extrato seco desengordurado	mín. de 8,5%	mín. de 8,4%	mín. de 8,2%	mín. de 8,3%	mín. de 8,4%
Índice crioscópico	não superior a -0,52 °C	-0,512 a -0,531 °C			

*EC, 1992; EC, 1997. **BRASIL, 2011. *** BRASIL, 1996.

O leite comercializado está sujeito a várias etapas de processamento ao longo da cadeia de abastecimento, tais como, ordenha, refrigeração, armazenamento a frio, homogeneização, tratamento térmico, envase e armazenamento. A homogeneização provoca modificações na estrutura física do leite, a qual consiste em uma passagem pressurizada em alta velocidade do leite através de uma abertura estreita, resultando em uma redução importante do tamanho dos glóbulos de gordura devido a força de cisalhamento, da cavitação e da turbulência, com consequente aumento da estabilidade da emulsão do leite (MICHALSKI & JANUEL, 2006; THIEBAUD *et al.*, 2003)..

A operação de homogeneização reduz os glóbulos de gorduras em média de 1-8 μm para 0,3-0,8 μm , o que previne a formação de nata e a coalescência durante o armazenamento em longo prazo. Geralmente, a homogeneização é realizada em pressões moderadas e sob aquecimento de 60-70 °C, embora as condições de pressão e temperatura possam variar com o tipo de equipamento e técnicas utilizados na operação (MICHALSKI & JANUEL, 2006; THIEBAUD *et al.*, 2003). A homogeneização também proporciona uma melhor digestibilidade do leite e um aumento da biodisponibilidade da membrana do glóbulo de gordura do leite e outros componentes que promovem a saúde (QI *et al.* 2015).

O tratamento térmico é o método mais tradicionalmente aplicado para aumentar a vida útil do leite, o qual pode ser aplicado em diferentes intensidades dependendo da extensão do prazo de validade desejável e, além disso, proporcionar uma ampliação da validade de algumas semanas sob refrigeração até 6 meses à temperatura ambiente. O processamento térmico do leite compreende os métodos de pasteurização por alta temperatura/curto tempo (*High temperature short time* – HTST) aplicando temperatura de 72 °C por 15-20 segundos, pasteurização alta empregando temperaturas de 80-90 °C por 15 segundos e ultra alta temperatura (*Ultra high temperature* – UHT) adotando temperaturas de 135-150 °C por 1-4 segundos (AMADOR-ESPEJO *et al.*, 2015).

A complexa composição do leite torna-o bastante perecível, por isso o tratamento térmico é usado para preservar a qualidade, aumentar a vida de prateleira e também destruir microrganismos deteriorantes e patogênicos, se presentes. Outras consequências promovidas pelo processamento térmico do leite incluem melhorias nas propriedades sensoriais, alterações irreversíveis na estrutura das proteínas do leite, formação de agregados proteicos, interações entre micelas de caseína, alterações que podem ser desejáveis ou prejudiciais dependendo das condições de temperatura do tratamento (RAIKOS, 2010; TUNICK *et al.*, 2016).

Efeitos adicionais do tratamento térmico no leite abrangem degradação da lactose em ácidos orgânicos e formação de lactulose (um dissacarídeo sintético produzido a partir da isomerização da lactose em meio alcalino durante o aquecimento ou por uma reação secundária na hidrólise da lactose). Além da destruição de algumas vitaminas e enzimas, hidrólise de proteínas e lipídios, alteração do equilíbrio cálcio-fósforo, efeitos relacionados ao sabor indesejável de cozido e perdas nutricionais causadas pela reação de Maillard (escurecimento não enzimático), que continua durante o armazenamento do leite tratado termicamente originando os compostos furosina, lactulose-lisina, hidroximetilfurfural

(HMF) e carboximetil-lisina (CML) (SAKKAS *et al.*, 2014). Por ser um carboidrato redutor, a lactose encontra-se bastante suscetível a participar da reação de Maillard com grupos amino livres de proteínas, peptídeos e aminoácidos livres, que ocorrem durante o aquecimento e o armazenamento (HUPPERTZ & GAZI, 2016).

O leite resultante do tratamento térmico é comercializado como leite UHT. Segundo a legislação brasileira, o leite UHT (*Ultra High Temperature*, ultra-alta temperatura) é definido como o leite homogeneizado que foi submetido durante 2 a 4 segundos a uma temperatura entre 130 °C, mediante um processo técnico de fluxo contínuo, imediatamente resfriado a menos de 32 °C e envasado sob condições assépticas em embalagens estéreis e hermeticamente fechadas. Assim, todo leite UHT integral, semi-desnatado ou parcialmente desnatado, acrescentadas as expressões “longa vida” e/ou “homogeneizado” comercializados seguem os padrões de identidade e qualidade fixados (BRASIL, 1996).

4.2. Lactose: aspectos digestivos, bioquímicos e tecnológicos

A intolerância ao leite é o termo geral comumente usado para descrever uma série de sintomas que inclui doenças mais específicas como a intolerância à lactose, a intolerância à proteína do leite de vaca e a alergia à proteína do leite de vaca. Nas reações adversas que envolvem as proteínas do leite, os sintomas de intolerância à lactose ou a presença de manifestações isoladas no trato gastrointestinal não são incluídos (WILSON, 2005).

A intolerância à lactose é uma condição de incapacidade de digestão do açúcar do leite consequente de uma deficiência na produção da enzima intestinal lactase em relação à quantidade de lactose ingerida (BAHNA, 1996). A causa mais frequente de má absorção de lactose é a diminuição progressiva da expressão da enzima durante a infância, condição definida como lactase não persistente (MISSELWITZ *et al.*, 2013).

A baixa atividade da enzima lactase pode resultar em má digestão de lactose e sintomas de desconforto gastrointestinal, categorizados como intolerância à lactose, se o consumo desse dissacarídeo for maior que a capacidade do trato gastrointestinal em digeri-

lo. Além de ser bastante influenciada por fatores como dose ingerida, tempo de trânsito intestinal, consumo de enzima β -galactosidase exógena e consumo habitual de produtos lácteos para adaptação da microbiota intestinal (BROWN-ESTERS *et al.*, 2012; DENG *et al.*, 2015).

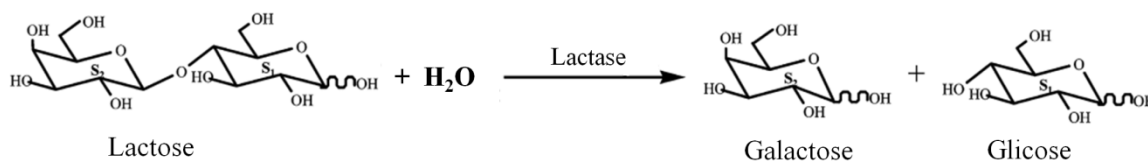
Os sintomas de intolerância à lactose geralmente não ocorrem até que a atividade enzimática intestinal esteja reduzida a menos de 50%. A fermentação da lactose não absorvida presente nos intestinos delgado e grosso pela microbiota colônica produz ácidos graxos de cadeia curta e os gases hidrogênio, metano e dióxido de carbono, que em excesso provocam os sintomas de dor abdominal, distensão, flatulência e cólicas resultantes do aumento do tempo de trânsito intestinal e da pressão intracolônica, nos indivíduos suscetíveis (BROWN-ESTERS *et al.*, 2012; DENG *et al.*, 2015).

A lactose mal absorvida também aumenta a carga osmótica e arrasta eletrólitos e água para o lúmen, provocando um rápido trânsito intestinal e conseqüentemente diarreia. A maioria das pessoas com baixa atividade da enzima lactase toleram pequenas quantidades de lactose sem apresentar sintomas, menos de 12 gramas, especialmente quando acompanhadas de outros alimentos ou fracionada ao longo do dia (BROWN-ESTERS *et al.*, 2012; DENG *et al.*, 2015).

Intolerantes à lactose estão teoricamente mais predispostos a uma inadequada mineralização óssea devido à menor ingestão de cálcio resultante da diminuição do consumo de leite e derivados, um problema reconhecido em muitos distúrbios que afetam pacientes pediátricos. Dessa maneira, a intolerância à lactose é um problema complexo de importância para a saúde pública e individual, requerendo maiores investigações (LULE *et al.*, 2016).

Os produtos lácteos com lactose hidrolisada, produzidos pela quebra desse carboidrato por enzimas β -galactosidases, têm conquistado cada vez mais espaço, principalmente pelo crescente conhecimento dos problemas causados pela intolerância à lactose (JELEN & TOSSAVAINEN, 2003). Embora existam algumas abordagens tecnologicamente viáveis para a produção de leite e produtos lácteos com lactose hidrolisada, o recurso mais habitual e conveniente para a indústria de alimentos compreende a hidrólise enzimática da molécula de lactose em glicose e galactose (Figura 1), que resulta um sabor mais doce e preço final do produto elevado relativo aos custos do procedimento enzimático (JELEN & TOSSAVAINEN, 2003).

Figura 1. Expressão da hidrólise enzimática da molécula de lactose.



Fonte: adaptado de SRIVASTAVA *et al.*, 2014. p. 7269.

A lactase (β -D-galactosidase ou β -D-galactosídeo-galactohidrolase, E.C. 3.2.1.23) é amplamente distribuída na natureza e pode ser isolada de diferentes fontes, como plantas (amêndoas, pêssegos, damascos, maçãs), órgãos animais, leveduras, bactérias e fungos. Quando se utiliza apenas uma enzima para a hidrólise da lactose, vários fatores devem ser considerados, tais como a concentração, o pH, a temperatura máxima admissível e o tempo de contato, a atividade enzimática e o custo. Um tempo de contato extensivo a 35-45 °C pode ser necessário para reduzir os custos, mas isso geralmente resulta em grande crescimento microbiano no leite cru (HARJU *et al.*, 2012).

Em produtos tratados por calor, um aumento dos açúcares redutores, glicose e galactose obtidas depois da conversão da lactose, provocam reações de Maillard mais intensas, devido a maior reatividade desses monossacarídeos na reação em comparação ao dissacarídeo lactose. No entanto, a intensidade da reação de Maillard durante o tratamento de calor pode ser reduzida significativamente quando a adição da enzima ocorre em condições assépticas após o tratamento térmico, o que reduz os níveis de escurecimento, embora isso não diminua a velocidade da reação de Maillard durante o armazenamento, devendo a temperatura durante a estocagem ser mantida a mais baixa possível (HARJU *et al.*, 2012).

A lactase utilizada deve ser produzida por microrganismos geralmente reconhecidos como seguros (GRAS). A qualidade da lactase utilizada pode ter um impacto significativo sobre a qualidade sensorial, algumas são conhecidas por conterem atividades colaterais que podem afetar o sabor. Além disso, a hidrólise de 70% de lactose no leite aumenta a doçura em uma quantidade correspondente a uma adição de cerca de 2% de sacarose (HARJU *et al.*, 2012).

A lactase ou β -galactosidasas também são utilizadas na indústria de alimentos para melhorar a doçura, a solubilidade e o sabor de produtos lácteos. As principais β -

galactosidases de origem microbiana utilizadas industrialmente são obtidas a partir de *Aspergillus* sp. e *Kluyveromyces* sp., e as enzimas obtidas a partir de *Kluyveromyces lactis* são mais amplamente empregadas. As enzimas derivadas das bactérias lácticas são geralmente reconhecidas como seguras, então elas podem ser utilizadas sem extensiva purificação e algumas cepas tem atividade probiótica (HUSAIN, 2010).

As β -galactosidases também são largamente distribuídas em tecidos de plantas, atuando em vários processos biológicos, tais como o crescimento das plantas, amadurecimento das frutas e na hidrólise da lactose (HUSAIN, 2010). Além disso, β -galactosidases recombinantes podem ser produzidas através da manipulação genética, selecionando propriedades singulares e adaptando-as a hospedeiros específicos que produzirão a enzima com níveis elevados de transgalactosilação sob condições determinadas (CHEN *et al.*, 2009; ANSARI & SATAR, 2012; NIE *et al.*, 2014).

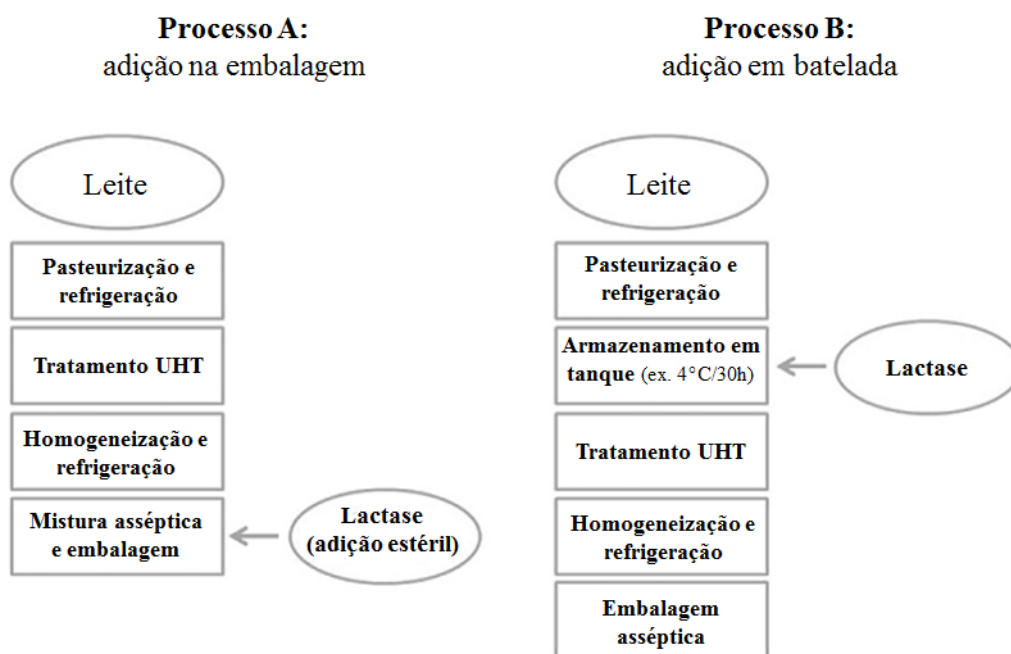
A hidrólise da lactose pode ser realizada por processos alternativos operando com hidrólise ácida, reatores de membrana, sistemas imobilizados e extratos celulares de *lactobacilli*, além da utilização de enzimas solúveis. A hidrólise ácida apresenta desvantagens relativas ao uso de pH em torno de 1,2 a temperaturas de 150 °C por um curto período de tempo no processo para viabilidade do método. No processo por reator de membrana, a complexidade do processo não o torna comercialmente atrativo. Em sistemas de enzima imobilizada, outras atividades da enzima nas células podem causar sabor indesejado no produto que podem não ser inativados. Assim, o uso de enzimas solúveis continua mais economicamente viável do que os investimentos em processos complexos, devido principalmente ao preço decrescente das lactases comerciais (HARJU *et al.*, 2012).

O método mais simples para produção de leite com lactose hidrolisada utilizando enzimas solúveis consiste na adição em batelada da lactase antes do tratamento térmico, empregando sistemas de batelada de uso único, sistemas de recuperação enzimática ou enzimas mobilizadas. Outra forma de processamento utiliza o aperfeiçoamento da tecnologia das embalagens assépticas com adição estéril da lactase em cada caixa de leite, o que reduz os custos operacionais por usar menores quantidades de enzima. O processo de adição na embalagem também viabiliza menor tempo de espera que a hidrólise em batelada e permite atingir concentrações mais baixas de lactose (TROISE *et al.*, 2016).

Ambos os processos compartilham as etapas comuns ao tratamento do leite, diferenciando-se na etapa de adição da lactase feita diretamente na embalagem ou em

batelada (Figura 2). O leite com lactose hidrolisada obtido pela adição em embalagem contém lactase ativa durante o período de validade e, conseqüentemente, a hidrólise da lactose continua durante o armazenamento. O leite com lactose hidrolisada obtido pela adição em batelada não contém atividade de lactase devido a inativação enzimática durante o tratamento térmico, uma vez que nesse processo a enzima é adicionada antes da aplicação de calor e a hidrólise é interrompida quando é atingido o final do período de incubação para ação da enzima em batelada (TROISE *et al.*, 2016).

Figura 2. Fluxograma dos processos tecnológicos utilizados para a produção de leite com lactose hidrolisada.



Fonte: adaptado de TROISE *et al.*, 2016. p. 515.

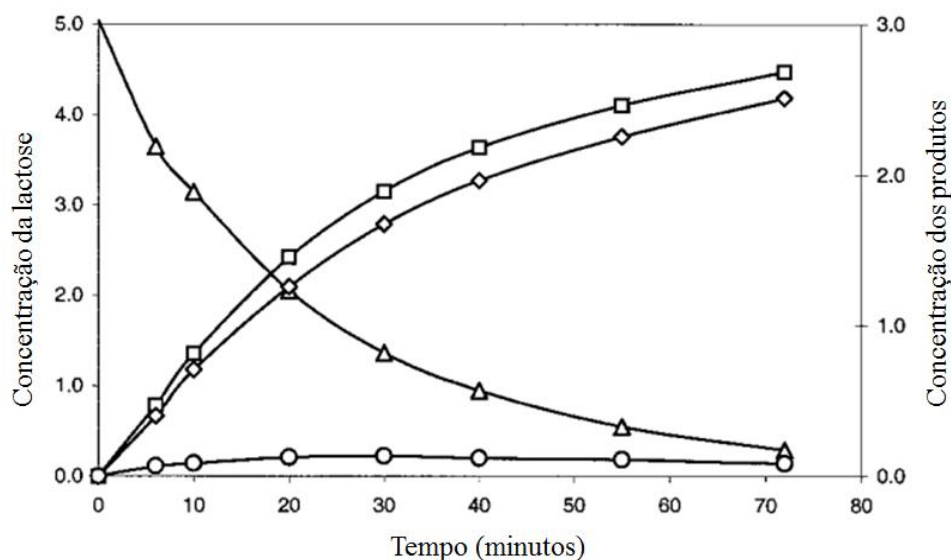
A presença de lactase ativa na embalagem final pode ter algumas desvantagens, uma vez que a elevada concentração de glicose e galactose torna o produto menos estável em comparação com o leite UHT convencional. De fato, sabe-se que a lactase comercial tem atividade enzimática paralela que poderia gerar outros intermediários reativos, provocando o aparecimento de sabor indesejado e gerando escurecimento não enzimático durante o período de validade (TROISE *et al.*, 2016).

A partir da hidrólise enzimática, a lactose produz principalmente glicose e galactose, embora alguns galacto-oligossacarídeos também sejam produzidos pela

atividade de transferase da enzima β -galactosidase (IWASAKI *et al.*, 1996) (Figura 3). O modelo cinético frequentemente utilizado para descrever a hidrólise enzimática da lactose é a equação de Michaelis-Menten com inibição competitiva pela galactose produzida. Os produtos são simultaneamente formados sob o efeito do complexo enzima-substrato até atingir o estado de equilíbrio (ZHOU *et al.*, 2003).

O modelo pressupõe que a molécula de glicose é a primeira a ser liberada do sítio ativo, deixando um complexo enzima-galactosil, que em seguida é transferido para um receptor contendo um grupo hidroxil e sofre a hidrólise. A formação de galacto-oligossacarídeos aumenta para uma concentração máxima, seguida por uma diminuição lenta (ZHOU *et al.*, 2003).

Figura 3. Reação típica da hidrólise enzimática da lactose com concentração inicial de 5% (m/v) por enzima imobilizada. (Δ) lactose, (\diamond) galactose, (\square) glicose, (\circ) galacto-oligossacarídeo.



Fonte: adaptado de ZHOU *et al.*, 2003. p. 130.

A demanda por produtos sem lactose pelos consumidores intolerantes criou um segmento de mercado para esse gênero alimentício. Leite e produtos lácteos sem lactose podem conter quantidades variáveis de lactose residual, o que é de extremo interesse para o bem-estar do consumidor, pois o limiar de tolerância à lactose é variável entre os indivíduos. Discute-se, ainda, como será definido o teor de lactose dos produtos

alimentares sem lactose, sendo proposto pelo grupo de trabalho *Issues of Nutrition of the German Society of Food Chemistry (LChG)* três categorias para declaração em alimentos: baixo teor de lactose para produtos alimentares com menos de 1% de lactose, teor muito baixo de lactose para um teor menor que 0,1% e livre de lactose para produtos com concentração menor que 0,01% de lactose e seus produtos de degradação (MORLOCK *et al.*, 2014).

A legislação brasileira considera a concentração de lactose mínima de 4,3% no leite (BRASIL, 1952). Bem como, estabelece que os alimentos cujo teor original de lactose tenha sido alterado devem declarar no rótulo a concentração de lactose remanescente (BRASIL, 2016). Outro regulamento, referente aos alimentos para dietas com restrição de lactose, estabelece que alimentos para dietas com restrição de lactose são classificados como isentos de lactose, quando a quantidade de lactose é igual ou menor que 0,1 grama por 100 gramas ou 100 mililitros do produto final a ser consumido, e como baixo teor de lactose, quando o conteúdo de lactose é igual ou menor que 1 grama por 100 gramas ou 100 mililitros do produto final. Além de definir que os valores de lactose e galactose devem ser declarados em gramas na tabela de informação nutricional (BRASIL, 2017).

O interesse da indústria de alimentos em suprir a demanda dos indivíduos com baixa atividade de lactase através da oferta de leite sem lactose tem aumentado progressivamente. Porém, a caracterização qualitativa, tecnológica e nutricional do leite modificado, assim como a avaliação da qualidade e do processamento, recebera pouca atenção. O que remete maiores cuidados, pois o leite com lactose hidrolisada apresenta elevada reatividade para a reação de Maillard e estabilidade química mais limitada quando armazenado a 20 °C (MESSIA *et al.*, 2007).

Os métodos analíticos disponíveis para a análise de lactose incluem polarimetria (método 896.01-1896, AOAC, 2005), detecção por infravermelho médio (método 972.16-1972, AOAC, 2005), fotometria (método DNS [MILLER, 1959], método Somogyi-Nelson [NELSON, 1944], método Fenol-Sulfúrico [DUBOIS *et al.*, 1956]), gravimetria (método Musson-Walker [HAMMOND, 1948]) e titulometria (método Lane-Eynon [LANE & EYNON, 1923]; método Luff-Schoorl [SHOORL, 1929]). A maioria desses métodos não permite a diferenciação individual entre carboidratos e não são adequados para identificar a lactose em produtos alimentares contendo diferentes açúcares (MORLOCK *et al.*, 2014; SRIVASTAVA *et al.*, 2014). Por isso, o desenvolvimento de métodos para análise de leite

que seja capaz de distinguir os analitos em estudo torna-se essenciais para a indústria de alimentos.

4.3. Espectroscopia no infravermelho próximo (NIR)

A região do infravermelho próximo (*Near Infrared* – NIR) foi identificada por Herschel no início do século XIX e avanços na capacidade computacional para manipulação matemática da informação química codificada no espectro e na sensibilidade dos detectores para captar a baixa energia da radiação NIR foram necessários a fim de que o potencial analítico da região NIR fosse assegurado (SCOTTER, 1990).

A região eletromagnética do espectro de infravermelho próximo compreende os comprimentos de onda de 800 a 2500 nm. As bandas de absorção são originadas a partir de sobretons e combinações das vibrações fundamentais que ocorrem na região do infravermelho médio, comumente entre 3000 e 20000 nm (WILLIAMS & STEVENSEN, 1990). A energia da radiação NIR e as implicações das vibrações de combinação permitem o fornecimento de informações estruturais complexas que fazem o método ter capacidade de utilização em determinações dos constituintes da matéria orgânica em vários tipos de amostras (COZZOLINO, 2015).

As análises usando espectroscopia no infravermelho próximo (*Near Infrared Spectroscopy* – NIRS) dependem da lei de Lambert-Beer, que mostra que a absorbância obtida está diretamente relacionada com a concentração do analito na amostra. O espectro de absorção contém informação relativa a quantidade e a forma na qual o hidrogênio está ligado a outra espécie atômica em um agrupamento H-X, onde o X é qualquer átomo com massa maior do que 1 (SCOTTER, 1990).

A espectroscopia NIR é um tipo de espectroscopia vibracional que emprega a energia do fóton em uma gama de energia maior do que o necessário para promover uma transição vibracional fundamental (entre dois níveis vibracionais ou seus harmônicos e combinações de duas ou mais vibrações) e inferior aos níveis típicos necessários para a excitação dos elétrons em moléculas (PASQUINI, 2003).

A combinação da energia de radiação com a diferença de energia entre dois níveis vibracionais provoca uma resposta seletiva do sistema molecular à radiação incidente, o que significa que em uma dada faixa de comprimento de onda algumas frequências serão absorvidas, outras que não correspondem a nenhuma das diferenças de energia possíveis para a molécula não serão absorvidas e algumas serão parcialmente absorvidas. Para que a vibração seja ativa, é necessário que o campo elétrico oscilante da onda eletromagnética possa interagir com a molécula e isso ocorre quando o deslocamento dos átomos em um modo vibracional produz uma mudança no momento dipolo da ligação (PASQUINI, 2003).

A intensidade de uma dada banda de absorção está associada com a magnitude da mudança do momento dipolo da ligação durante o deslocamento de átomos em uma vibração e o seu grau de anarmonicidade, e a intensidade de absorção em relação ao comprimento de onda constitui o espectro de absorção de uma substância. Os espectros NIR são dominados por sobretons e bandas de combinação que, associados a outros efeitos como acoplamento e ressonância entre diferentes vibrações do mesmo grupo funcional, contribuem para a complexidade dos espectros (PASQUINI, 2003).

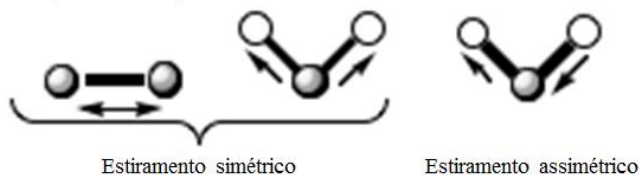
Os espectros moleculares resultam dos modos vibracionais dos núcleos atômicos das moléculas, realizando movimentos em que giram, vibram, se agitam e se dobram em relação aos seus centros de gravidade. A vibração e flexão das ligações moleculares exibem atividade espectroscópica vibracional com frequência específica que pode ser medida. A frequência da vibração não depende da amplitude, mas a simetria da molécula é um fator importante na determinação da frequência e da amplitude de uma vibração molecular. Portanto, a amplitude da absorção da energia NIR a qualquer comprimento de onda é determinada pela sua absortividade e pelo número de moléculas encontradas dentro da trajetória do feixe e supõe-se que uma mudança na resposta espectral está relacionada a uma concentração pela lei de Lambert-Beer (WORKMAN JR. & WEYER, 2012).

Existem diferentes tipos de vibrações moleculares que absorvem comprimentos de ondas específicos ou frequências de energias do infravermelho próximo dependendo do tipo de ligação (Figura 5). As vibrações de estiramento são uma variação do comprimento da ligação entre os átomos e podem ser simétricas (quando os átomos se movimentam na mesma direção) ou assimétricas (quando os átomos se movem em direções opostas). A flexão é uma alteração no ângulo da ligação entre os átomos, pode ser simétrica ou assimétrica (tesoura – *scissoring*, balanço – *rocking*) e ocorrer no plano ou fora do plano

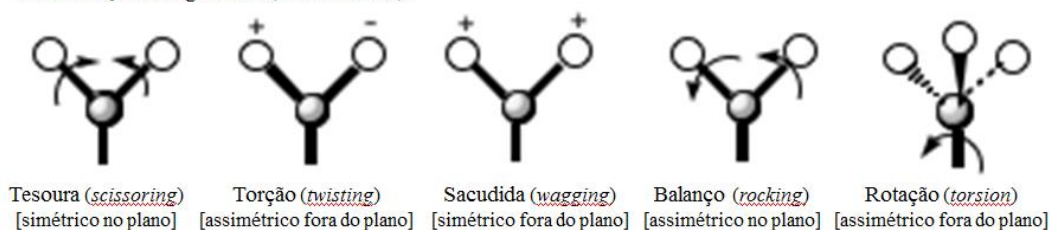
(sacudida – *wagging*, torção – *twisting*, rotação – *torsion*) (WORKMAN JR. & WEYER, 2012).

Figura 4. Tipos de vibrações moleculares.

Deformações axiais (estiramento):



Deformações angulares (dobramento):



Fonte: adaptado de COLUMBIA UNIVERSITY [online], 2007.

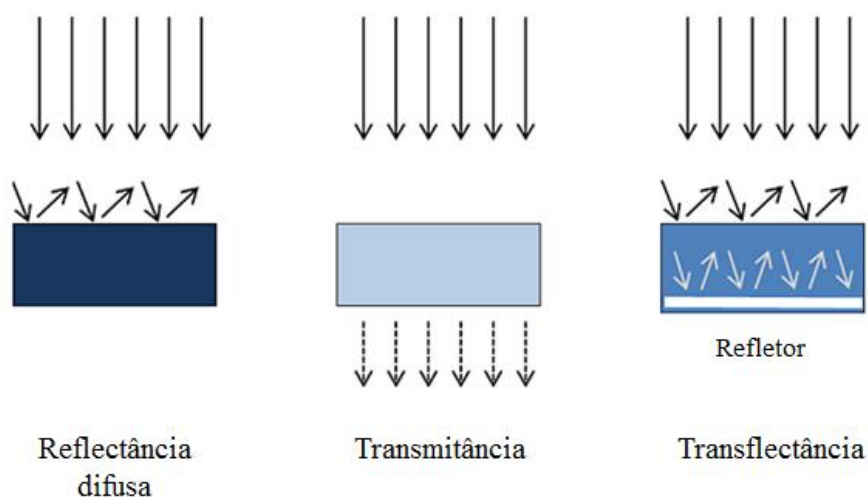
As ligações do tipo H-X são as mais ativas e intensas, uma vez que as massas de X e hidrogênio diferem significativamente e possuem harmônicas de suas vibrações de estiramento e flexão que absorvem energia ou vibram com uma frequência maior do que as bandas fundamentais. Assim, as moléculas ativas para a energia NIR consistem em modelos mecânicos com dipolos vibracionais, cada modelo de dipolo vibra com uma frequência e amplitude específicas (WORKMAN JR. & WEYER, 2012). Então, na região espectral do infravermelho próximo, as moléculas contendo grupos funcionais C-O, C=O, N-H e O-H são excitadas e passam a expressar vibrações de estiramento e deformação angular que produzem os sobretons e as combinações vibracionais fundamentais que podem ser medidas (HUCK, 2015).

Durante a aquisição de um espectro infravermelho, a amostra é irradiada e a radiação absorvida excita as vibrações moleculares. Se a radiação incidente for transmitida através da amostra, o modo de medida é chamado de transmitância. Caso a radiação incidente seja refletida pela amostra, o modo de medida é denominado de reflectância, que pode variar entre reflectância especular (quando a radiação é refletida a partir da superfície

da amostra) e reflectância difusa (quando a radiação penetra na amostra e é parcialmente refletida e transmitida) (NETCHACOVITCH *et al.*, 2015).

A transfectância representa um caso especial de medição por transmissão e reflexão utilizado para medições de líquidos ou soluções turvas, na qual o feixe luminoso de entrada passa através da amostra, em seguida é refletido sobre uma substância espelhada não absorvente e depois é direcionado ao detector após penetrar na amostra uma segunda vez (Figura 5). Nesse modo de medida o comprimento do percurso óptico é duplicado, pois o feixe de radiação passa duas vezes pela amostra (HUCK, 2015).

Figura 5. Tipos de medições para aquisição do espectro de infravermelho.



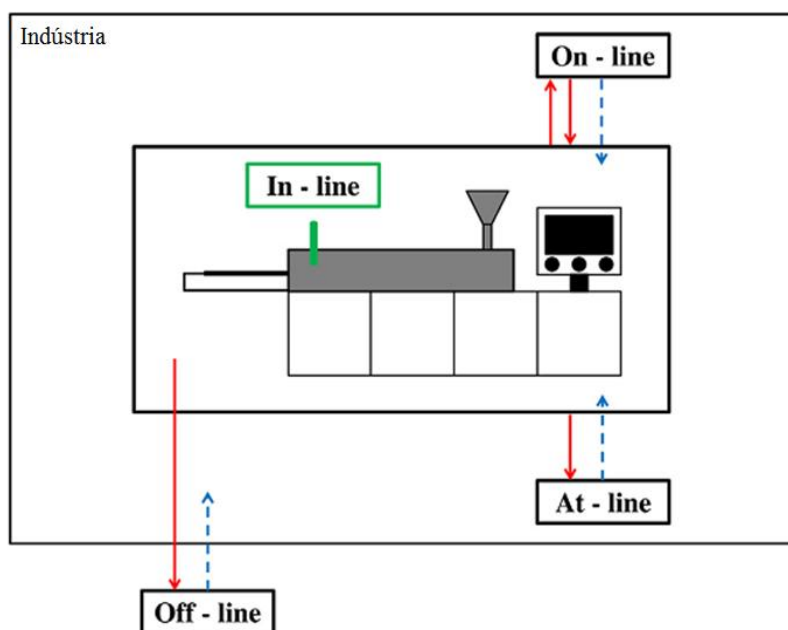
Fonte: adaptado de HUCK, 2015. p. 386.

A espectroscopia NIR apresenta características significativas, como rapidez (um minuto ou menos por amostra), ser não destrutiva e não invasiva, com alta penetração do feixe de radiação, adequação para uso em linha, aplicação quase universal (qualquer molécula contendo ligações CH, NH, SH ou OH) e com exigências mínimas de preparação da amostra, fatores que consolidam essa abordagem espectrofométrica (PASQUINI, 2003).

Outra vantagem interessante da espectroscopia NIR é a possibilidade de implementação em sistemas *off-line*, *at-line*, *on-line* ou *in-line* (Figura 6). No sistema *off-line*, a análise é realizada fora do fluxo de processo após a coleta da amostra final e frequentemente informações valiosas são perdidas. Quando a amostra é retirada e imediatamente analisada próxima ao processo, esse sistema é chamado *at-line*. No caso da

amostra ser retirada do fluxo de processo em análise e em seguida ser reintroduzida, o sistema é denominado *on-line*. Quando a amostra é mantida no fluxo do processo durante a análise, o sistema de medição é chamado *in-line*, e fornece informações em tempo real sobre o processamento da amostra (NETCHACOVITCH *et al.*, 2015).

Figura 6. Esquema das técnicas de monitoramento do processo de produção. Seta vermelha = transferência da amostra; seta tracejada azul = transferência de dados.



Fonte: adaptado de NETCHACOVITCH *et al.*, 2015. p. 24.

A precisão das medições espectroscópicas baseia-se, principalmente, nas especificações técnicas do instrumento e na apresentação da amostra. O tipo de fonte de luz e a tecnologia de detecção, configuração de medição, razão sinal-ruído, faixa de comprimento de onda e resolução estão entre os parâmetros técnicos mais relevantes. Além de acurácia e precisão, o instrumento deve oferecer um desempenho robusto em flutuações de iluminação, mudanças de umidade, variações de temperatura e ambientes com interferências vibratórias (COLLELL *et al.*, 2012).

Os instrumentos dispersivos baseados em redes de difração foram primariamente empregados na espectroscopia NIR e responsáveis pela consolidação desta como uma ferramenta analítica. Os espectrômetros baseados em monocromadores de grade apresentam a vantagem do custo relativamente baixo, quando comparados com outros

instrumentos de varredura que utilizam tecnologias modernas, e desempenho adequado para muitos métodos analíticos (PASQUINI, 2003).

As principais desvantagens dos espectrômetros dispersivos são varredura espectral mais demorada e uma menor precisão do comprimento de onda, que se deteriora com o uso devido ao desgaste mecânico do mecanismo e pode ser um ponto negativo em relação à manutenção de modelos multivariados. A presença de peças móveis constitui outro fator que limita o uso dos instrumentos dispersivos em campo e em ambientes mais agressivos. Em contrapartida, a evolução na tecnologia de produção dos sensores oferece à óptica dispersiva uma vida útil mais longa e com relação sinal-ruído melhorada (PASQUINI, 2003).

A espectroscopia no infravermelho próximo com transformada de Fourier (FT-NIR) é uma tecnologia que apresenta melhorias na reprodutibilidade espectral, na exatidão e na precisão da discriminação do comprimento de onda pelo uso de interferômetro. Comparada com a espectroscopia NIR dispersiva, a espectroscopia FT-NIR tem a vantagem de maior estabilidade do instrumento, profundidade de penetração da luz e capacidade preditiva (GASPARDO *et al.*, 2012).

A aplicação da esfera de integração no FT-NIR permite recuperar a fração de radiação não absorvida a partir da amostra e refletida na direção da superfície interna da esfera, retomando informações importantes que de outra forma poderiam ser perdidas. A espectroscopia FT-NIR proporciona um grande potencial para monitoramento on-line de controle de qualidade em indústrias de alimentos (GASPARDO *et al.*, 2012). Embora o FT-NIR ofereça maiores razão sinal-ruído, resolução de banda melhorada e determinação precisa e rápida de frequência, a tecnologia baseada em interferômetro o torna mais caro (PLANS *et al.*, 2013).

A miniaturização dos componentes da espectroscopia vibracional permite a utilização de instrumentos portáteis mais práticos e econômicos, além dos espectrômetros de bancada mais complexos e onerosos manipulados em laboratórios (AYVAZ & RODRIGUEZ-SAONA, 2015). A tecnologia portátil oferece versatilidade e robustez, reduzindo o tamanho e o peso dos sistemas de espectroscopia sem comprometer o desempenho e a reprodutibilidade (PLANS *et al.*, 2013).

As vantagens de sistemas de espectroscopia vibracional portáteis incluem o baixo custo, compactação, robustez, resistência, leveza, geralmente requerem um mínimo de

treinamento para operação do equipamento e podem ser usados para análises de rotina (PLANS *et al.*, 2013). Outra vantagem do equipamento portátil é a possibilidade de transporte do dispositivo e manejo *in situ* do parâmetro que será medido (CAYUELA & WEILAND, 2010).

Os instrumentos portáteis oferecidos comercialmente apresentam peso e tamanho que possibilitam o manuseio com apenas uma mão, empregam sistema microeletromecânico ou filtro linear variável sem parte móvel e são capazes de obter espectros dos compostos escaneados nas faixas de 1600 a 2400 nm ou 950 a 1650 nm. Esses tipos de sensores são altamente resistentes ao estresse mecânico e têm uso simplificado, o que representa o avanço na análise de amostras com acurácia, precisão e alta representatividade que permite levar o laboratório para a amostra, importante em rotinas de fiscalização, por exemplo (BASRI *et al.*, 2017; MODROÑO *et al.*, 2017).

As principais vantagens da utilização de técnicas espectroscópicas estão na aquisição rápida de dados da amostra, possibilidade de determinação simultânea de vários parâmetros de qualidade e capacidade para otimizar as análises custosas e demoradas (KAROUI & BAERDEMAEKER, 2007). A espectroscopia NIR associada à análise multivariada de dados tem potencial para a previsão quantitativa ou qualitativa de parâmetros físico-químicos, tais como, proteína, gordura e umidade em grãos, oleaginosas, carnes e produtos lácteos, sendo relatada como uma ferramenta rápida, consistente e econômica (NÚÑEZ-SÁNCHEZ *et al.*, 2016). Todos os avanços analíticos alcançados com relação à espectroscopia NIR permitem gerar dados químicos com grande quantidade de informação, ou seja, com muitas variáveis, das amostras de forma rápida.

4.4. Análise multivariada dos dados

A espectroscopia NIR funciona baseada na penetração da radiação na amostra, essa radiação é refletida e absorvida e/ou transmitida pelas ligações moleculares, resultando em mudanças na energia do feixe emitido que caracteriza algumas ligações químicas. Em particular, os dados obtidos a partir do infravermelho próximo normalmente contêm muita informação redundante e ruído, que pode ser minimizado e/ou eliminado por manipulações matemáticas conhecidas como pré-processamento (QU *et al.*, 2015).

A Quimiometria é uma área da química que aplica métodos matemáticos e estatísticos para selecionar as melhores condições experimentais e extrair o máximo de informação dos dados químicos. Os métodos quimiométricos comumente utilizados incluem planejamento e otimização de experimentos, métodos de reconhecimento de padrões e de calibração multivariada (ROGO *et al.*, 2007).

A combinação da espectroscopia NIR e análise multivariada dos dados é apropriada para presumir a composição química de muitos alimentos e produtos agrícolas com alta precisão. Para a análise multivariada é necessário o desenvolvimento do modelo matemático, que consiste em um processo complexo que requer a compreensão de um sistema constituído por amostra, instrumento, análise multivariada de dados e o analista (COZZOLINO *et al.*, 2011).

Outros elementos do sistema que devem ser considerados e compreendidos antes, durante e após o desenvolvimento do modelo são conhecimento e compreensão do método de referência, entendimento da base física e química do espectro, compreensão das interações entre a amostra e o instrumento, e a interpretação e validação do modelo para produzir um método confiável (COZZOLINO *et al.*, 2011).

A análise multivariada de dados é essencialmente classificada em métodos de regressão e classificação multivariada com base na complexidade dos dados estimados (KUMAR *et al.*, 2014). Os métodos de projeção multivariados tornam a visualização dos resultados com elevada dimensionalidade de variáveis obtidos pela espectroscopia NIR mais compreensível (RAJALAHTI & KVALHEIM, 2011).

Para o desenvolvimento de um modelo de classificação, as etapas de seleção das amostras, pré-processamento dos espectros, escolha de algoritmos para a seleção de variáveis, etapas de treinamento e previsão e a interpretação dos resultados obtidos devem ser consideradas com atenção (COZZOLINO *et al.*, 2011).

4.4.1. Preparação dos dados para análise

O pré-tratamento dos dados espectrais é uma etapa imprescindível antes da construção dos modelos multivariados. Os espectros NIR de reflectância ou transmitância

podem ser substancialmente influenciados por não linearidades introduzidas pela dispersão da luz, produzindo efeitos de espalhamento indesejados, como alterações aditivas e multiplicativas na linha de base que podem sobrepor a informação de interesse (ZEAITER *et al.*, 2005; RINNAN *et al.*, 2009).

A correção de linha de base consiste em técnicas de pré-tratamento aplicadas às linhas da matriz de dados, ou seja, às amostras e visam redução das variações aleatórias e de informações sistemáticas irrelevantes (FERREIRA, 2015a). As técnicas de pré-tratamento mais aplicadas ao espectro NIR para correção de linha de base são os métodos de correção de espalhamento e as derivadas espectrais (ZEAITER *et al.*, 2005; RINNAN *et al.*, 2009).

Entre os métodos de transformação para correção de espalhamento estão incluídos correção de espalhamento multiplicativo (*Multiplicative Scatter Correction* – MSC), variação normal padrão (*Standard Normal Variate* – SNV) e normalização. Esses métodos reduzem os efeitos multiplicativos produzidos pela dispersão, pelo tamanho das partículas e pelas modificações de multicolinearidade ao longo do espectro sem alterar o perfil original (ZEAITER *et al.*, 2005; RINNAN *et al.*, 2009). A correção espectral MSC é obtida a partir de uma regressão por mínimos quadrados de cada espectro contra um espectro de referência, que em geral corresponde ao espectro médio. Na correção SNV, cada espectro é centrado na média e dividido pelo seu respectivo desvio padrão (FERREIRA, 2015a).

As derivadas espectrais são representadas principalmente por derivadas de Norris-Williams (NW) e derivadas com filtros polinomiais de Savitzky-Golay (SG), utilizam a suavização dos espectros em conjunto ao cálculo da derivada com a finalidade de diminuir o efeito prejudicial na relação sinal-ruído que as derivadas finitas convencionais apresentariam. As derivadas têm a capacidade de remover os efeitos aditivos e multiplicativos nos espectros e deslocar o espectro resultante para o zero de absorbância (ZEAITER *et al.*, 2005; RINNAN *et al.*, 2009).

A primeira derivada corrige os efeitos na linha de base e a segunda derivada corrige os efeitos na linha de base e a tendência de inclinação linear. As derivadas provocam aumento do ruído espectral e desta forma devem ser utilizadas junto a filtros de suavização. O filtro de suavização SG usa uma suavização de janela simétrica, isto é, o número de pontos de dados em cada lado do ponto central é igual; e quanto maior for o

número de pontos dessa janela, maior será a redução do ruído, mas também maior será a distorção do sinal, portanto, é de grande importância selecionar cuidadosamente o número de pontos para essa janela (ZEAITER *et al.*, 2005; RINNAN *et al.*, 2009).

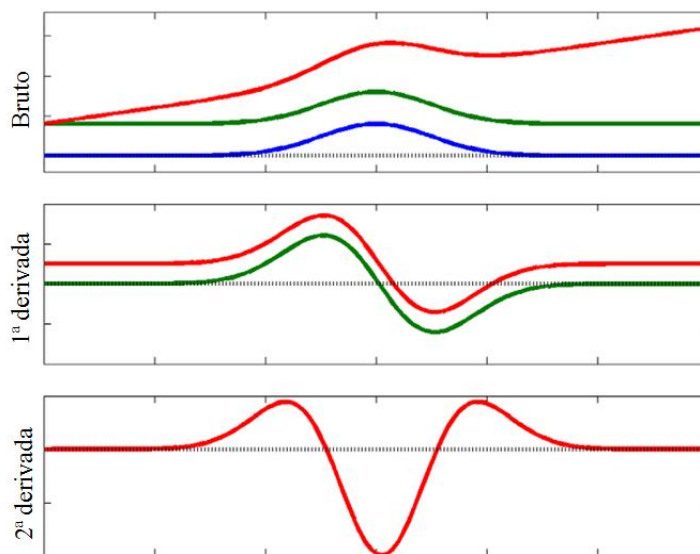
O método desenvolvido por Savitzky & Golay (1964) calcula a suavização e diferenciação dos dados por uma técnica de quadrados mínimos, na qual o valor dos quadrados mínimos de um dado ponto é calculado como uma combinação ponderada de si mesmo e m pontos em cada lado da matriz de dados, com os pesos obtidos por ajuste polinomial. Isso corresponde à realização de um ajuste de quadrados mínimos ($2m + 1$ ponto) em uma janela móvel de dados (GORRY, 1990).

Na derivada SG, polinômios (tipicamente de ordem 2 a 8) são ajustados localmente a cada um dos pontos centrais do conjunto de dados espectrais na janela juntamente com um número ímpar de seus pontos vizinho selecionados, tipicamente 5 a 11 ou mais pontos, de cada lado. O polinômio ajustado é então diferenciado analiticamente para gerar a derivada em cada ponto central, essa diferenciação pode ser realizada quantas vezes forem necessárias para gerar a ordem desejada da derivada e todas as janelas tenham sido alisadas. Na prática, a amplificação do ruído geralmente atinge um nível inaceitável na terceira e quarta derivadas (YEOW *et al.*, 2005).

A abordagem de Savitzky-Golay apresenta como principal desvantagem a supressão de dados por m pontos em cada extremidade. Isso ocorre porque a técnica requer m pontos à direita e à esquerda de um ponto central para calcular o valor de quadrados mínimos exigido, isso se torna menos importante quando a matriz de dados é composta por um grande número de comprimentos de onda (GORRY, 1990).

Vale ressaltar que a derivada de uma função em um ponto qualquer da curva é a inclinação da reta tangente à curva naquele ponto, assim, a derivada de uma banda terá dois lobos: um positivo, relativo ao lado esquerdo, e outro negativo, relativo ao lado direito da banda. Os máximos, positivo e negativo da derivada, correspondem aos pontos de inflexão da banda original (Figura 7). A relação entre a área abaixo de uma banda do espectro e a concentração é preservada quando se toma a derivada, isto é, os detalhes espectrais não são alterados e a correspondência quantitativa entre o espectro e a concentração permanece válida (FERREIRA, 2015a).

Figura 7. Correção da linha de base pela aplicação da 1ª e 2ª derivadas. A 1ª derivada corrige os deslocamentos da linha de base. A 2ª derivada corrige tanto o deslocamento, como a inclinação constante na linha de base.



Fonte: adaptado de RINNAN *et al.*, 2009. p. 1214.

Entre as técnicas de pré-tratamento de dados aplicadas às variáveis do espectro NIR destaca-se a centralização na média. Essa técnica consiste em uma translação de eixos e o valor médio de cada variável é calculado e subtraído de cada valor da variável, o que torna a média de todas as variáveis igual a zero. A consequência prática da centralização na média dos dados é um modelo mais simples e interpretável, e mais informações podem ser descritas (BOYSWORTH & BOOKSH, 2007).

O pré-tratamento dos dados é uma parte essencial da análise multivariada dos dados ao remover as variações não desejadas e assim focar na resposta de interesse. Um pré-tratamento adequado influencia diretamente o resultado em todas as etapas seguintes e, portanto, o sucesso de toda a experiência. A escolha do pré-tratamento dos dados requer uma estratégia criteriosa para que as conclusões obtidas sejam robustas. As abordagens de seleção de pré-tratamentos mais comumente usados são: tentativa e erro, no qual se aplicam diferentes estratégias de pré-tratamentos e seleciona-se o melhor desempenho de acordo com o objetivo da análise dos dados; inspeção visual, em que para cada estratégia de pré-tratamento de interesse, os dados pré-tratados são inspecionados visualmente para verificação da persistência de variações indesejáveis visíveis; e avaliação dos dados pré-

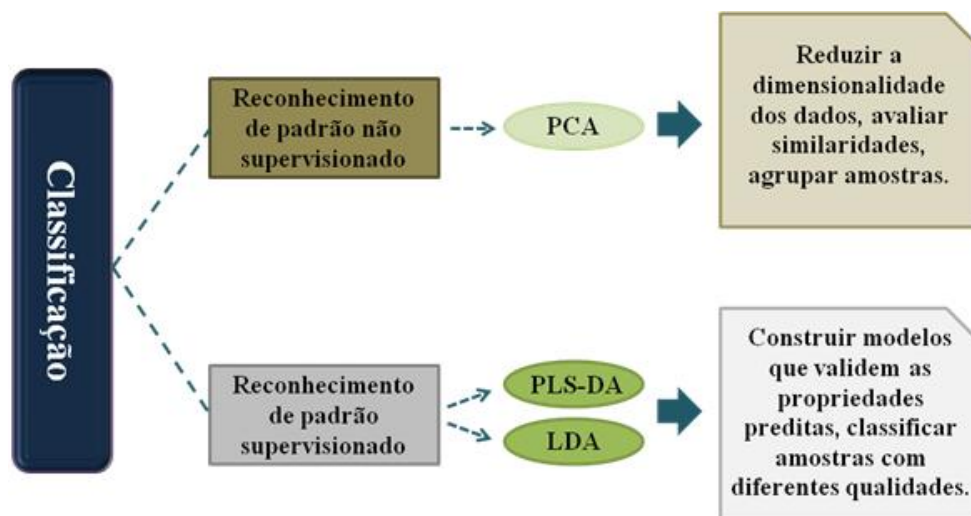
tratados por parâmetros de qualidade, que visam quantificar a presença das variações nos dados (ENGEL *et al.*, 2013).

A aplicação do pré-tratamento adequado elimina os efeitos indesejados e melhoram a eficiência dos modelos ao reduzir a variabilidade não modelada nos dados sem remover a informação procurada nos espectros com a finalidade da realização bem sucedida dos propósitos da análise de dados, sejam eles exploração, previsão ou classificação (ZEAITER *et al.*, 2005; RINNAN *et al.*, 2009; ENGEL *et al.*, 2013).

4.4.2. Reconhecimento de padrão

Os métodos de reconhecimento de padrão das amostras são classificados em métodos de reconhecimento de padrões supervisionados ou não supervisionados (Figura 8). Nos métodos supervisionados é necessário um conhecimento prévio das categorias que cada amostra pertence de modo a classificar novas amostras numa das classes conhecidas com base no seu padrão de medições, assim, o modelo de classificação é desenvolvido em um conjunto de treinamento de amostras com categorias pré-definidas. Entre os métodos clássicos para a classificação supervisionada estão o SIMCA (*Soft-Independent Modeling of Class Analogy*), a análise discriminante linear (*Linear Discriminant Analysis – LDA*) e a análise discriminante pelo método dos quadrados mínimos parciais (*Partial Least Squares Discriminant Analysis – PLS-DA*). Nos métodos não supervisionados, as amostras são agrupadas sem conhecimento prévio a respeito da classificação das amostras. A análise de componentes principais (*Principal Component Analysis – PCA*), por exemplo, é utilizada para esse propósito, principalmente para a análise exploratória dos dados (ROGGO *et al.*, 2007; HUANG *et al.*, 2016).

Figura 8. Esquema dos métodos de reconhecimento de padrão supervisionado e não supervisionado.



Fonte: Adaptado de HUANG *et al.*, 2016. p. 29.

A análise de componentes principais (*Principal Component Analysis* – PCA) é o método de projeção de componentes principais (ou autovetores) habitualmente utilizado. Essa análise é usada como uma ferramenta para triagem, extração e compressão dos dados multivariados, cuja ideia principal consiste em reduzir a dimensão de um conjunto de dados composto por um grande número de variáveis correlacionadas e transformá-lo em um novo conjunto de variáveis não correlacionadas e ortogonais entre si, chamados componentes principais (*Principal Components* – PC), mantendo o máximo possível da variação presente no conjunto de dados original. Na matriz de dados, as linhas correspondem às amostras e as colunas representam as variáveis (COZZOLINO *et al.*, 2011; FERREIRA, 2015b; KUMAR *et al.*, 2014; RAJALAHTI & KVALHEIM, 2011).

Em cada PC, as amostras recebem um valor de escore que são representados no gráfico de *scores* e revelam padrões como aglomeração (*clusters*), tendência e amostras anômalas (*outliers*) nos dados. Da mesma maneira, cada variável recebe um valor de peso (*loading*) em cada PC que são representadas no gráfico de *loadings* e apontam variâncias entre as variáveis que podem ser usadas para interpretar os padrões observados no gráfico de *scores*. Em princípio, a PCA é uma técnica de visualização de dados que representa as amostras em um novo sistema de eixos, no qual a natureza multivariada dos dados pode ser visualizada em poucas dimensões descrevendo quase toda a informação contida nos dados originais, uma vez que as redundâncias são removidas sem alteração nas relações entre as

amostras (COZZOLINO *et al.*, 2011; FERREIRA, 2015b; KUMAR *et al.*, 2014; RAJALAHTI & KVALHEIM, 2011).

A análise discriminante pelo método dos quadrados mínimos parciais (*Partial Least Squares Discriminant Analysis* – PLS-DA) é um dos métodos supervisionados aplicados em classificação. PLS-DA foi desenvolvida a partir dos algoritmos de mínimos quadrados parciais (*Partial Least Squares* – PLS) sendo usada na classificação multivariada para modelar a relação entre duas matrizes, na qual uma relação linear entre a variável dependente (Y) e a variável independente (X) é estabelecida. Cada matriz contém informações diferentes sobre o mesmo grupo de amostras, a matriz Y contém informações pré-determinadas hipotéticas sobre as classes das amostras e a matriz X contém as respostas instrumentais. Assim, o desenvolvimento do modelo ocorre com base na análise de componentes principais, onde as matrizes X e Y são decompostas no produto de duas matrizes, escores e pesos, com uma ligeira rotação no eixo da componente principal que procura a covariância máxima de X com Y. As componentes principais são chamadas de variáveis latentes (*Latent Variables* – LV) e são combinações lineares das variáveis iniciais selecionadas que maximizam a discriminação entre os grupos (BORRÀS *et al.*, 2014; MARQUETTI *et al.*, 2016).

A matriz Y contém informações sobre as diferentes classes de amostras, assumindo valores hipotéticos como “0” ou “1” dependendo das classes representadas na coluna, e a modelagem consiste em treinamento e previsão. No treinamento as características dos dados são investigadas para encontrar um modelo para o comportamento das amostras e na previsão os dados que não participaram na etapa de treinamento são usados para avaliar a adequação do modelo construído. O valor limítrofe para a separação das classes é baseado no Teorema de Bayes, o qual assume que os valores dos vetores y previstos seguem uma distribuição semelhante à que será observada em amostras futuras (BORRÀS *et al.*, 2014; MARQUETTI *et al.*, 2016). Em síntese, a PLS-DA busca encontrar as variáveis e direções no espaço multivariado que discriminam as classes estabelecidas no conjunto de treinamento (HUANG *et al.*, 2016).

A análise discriminante linear (*Linear Discriminant Analysis* – LDA) é outro método de reconhecimento de padrão supervisionado bastante utilizado. LDA baseia-se na determinação de funções discriminantes lineares que maximizam a variância entre classes e minimizam a razão de variação dentro da classe. Em LDA, as classes devem seguir uma distribuição normal multivariada e serem linearmente separadas, além da capacidade de

agrupamento das matrizes de variância-covariância das classes. A LDA determina um hiperplano de menor dimensão sobre o qual os pontos serão projetados a partir da dimensão maior, selecionando uma direção que alcança a máxima separação entre as classes dadas. A variável latente obtida em LDA é uma combinação linear das variáveis originais, também chamada de variação canônica (BERRUETA *et al.*, 2007; STANIMIROVA *et al.*, 2010; HUANG *et al.*, 2016).

A LDA utiliza a distância de Mahalanobis para classificar as amostras desconhecidas e é amplamente empregada devido a sua simplicidade e menor suscetibilidade de sobrecarregar os dados de treinamento, o que propicia robustez (MOREIRA *et al.*, 2009; REINHOLDS *et al.*, 2015). Uma desvantagem da LDA é ser apropriado somente para dados em que o número total de objetos (amostras) é consideravelmente maior do que o número de variáveis (resposta instrumental). Quando essa condição não é cumprida, a compressão de dados por meio de PCA ou outro método de seleção de variáveis devem ser considerados (BERRUETA *et al.*, 2007; STANIMIROVA *et al.*, 2010; HUANG *et al.*, 2016).

O desempenho dos modelos de classificação é avaliado usando parâmetros como sensibilidade e especificidade. A sensibilidade é a habilidade do modelo para classificar corretamente as amostras, relacionando as amostras previstas para uma classe com as amostras que foram categorizadas nessa classe. A especificidade relaciona as amostras previstas como não pertencentes a uma classe com as amostras que não estão realmente nesta classe (BORRÀS *et al.*, 2014; MARQUETTI *et al.*, 2016).

Seleção de variáveis informativas para o modelo de classificação

A seleção de variáveis é uma etapa importante para melhorar o efeito de previsão do modelo ao selecionar bandas com melhor relação sinal-ruído relacionadas à resposta de interesse. A seleção dessas variáveis visa identificar um subconjunto de variáveis que produzem o menor erro possível e fornecem preditores mais rápidos e econômicos (ZEAITER *et al.*, 2005; CHEN *et al.*, 2011). A influência das regiões espectrais que não contêm informações críticas pode influenciar a capacidade preditiva dos modelos, uma vez que nem todas as variáveis ou suas regiões são igualmente interessantes para a modelagem,

por exemplo, as áreas de ruído e sinal saturado são bastante prejudiciais. A remoção das variáveis, nas quais o ruído domina sobre as informações relevantes, e a seleção de faixas espectrais leva a uma melhor precisão e desempenho dos modelos de classificação (XIAOBO *et al.*, 2010).

A seleção de variáveis baseia-se no princípio de seleção criteriosa de um número menor de comprimentos de onda relevantes selecionados a partir dos espectros originais. Várias abordagens para seleção de variáveis estão disponíveis, incluindo as abordagens univariadas que selecionam as variáveis que têm maior correlação com a resposta, as abordagens sequenciais que classificam as variáveis em ordem e as agrupam em progressões, e as abordagens multivariadas que aplicam seleção de variáveis tais como a eliminação de variáveis não informativas (*Uninformative Variable Elimination – UVE*), quadrados mínimos parciais por intervalos (*Interval Partial Least-Squares Regression – iPLS*), testes dos parâmetros de variáveis do modelo, algoritmo genético (*Genetic Algorithm – GA*) e algoritmo das projeções sucessivas (*Successive Projection Algorithm – SPA*) (XIAOBO *et al.*, 2010). Uma seleção de variáveis bem executada resulta em modelos mais simples com melhor capacidade de previsão em determinações quantitativas e em análises discriminantes, facilitando a interpretação (BALABIN & SMIRNOV, 2011).

Algoritmo genético (GA) e algoritmo das projeções sucessivas (SPA) são técnicas de seleção de variáveis úteis em análises espectroscópicas. GA é um método de busca estocástico ou aleatório e uma técnica popular de otimização heurística que emprega um processo probabilístico não local de busca baseada no princípio do sistema genético de seleção natural. O algoritmo genético segue cinco passos básicos: codificação de todas as variáveis, iniciação da população, avaliação das respostas, reproduções e mutações. As etapas de três a cinco são alternadas até que, baseado na melhor resposta obtida ou simplesmente em um número máximo de gerações, ou no tempo total permitido para a elaboração do modelo, um critério de terminação seja alcançado (CHENG *et al.*, 2016).

O algoritmo genético gera soluções úteis para problemas de otimização e busca baseada na codificação de subconjuntos de variáveis na forma de cadeias de valores binários (0/1) denominados cromossomos. Cada gene no cromossomo está associado a uma das variáveis disponíveis para seleção. Quando o valor do gene é igual a 1, os comprimentos de onda são incluídos no conjunto de dados para produzir o modelo de classificação e a evolução geralmente inicia a partir de uma população de indivíduos gerados aleatoriamente, evoluindo através de gerações. A aptidão de cada indivíduo na

população é avaliada em cada geração e múltiplos indivíduos são estocasticamente selecionados a partir da população atual (com base em sua aptidão) e modificados (recombinados e modificados aleatoriamente) para formar uma nova população (SILVA *et al.*, 2013).

Em contrapartida, SPA visa resolver o problema de colinearidade com redundância mínima, baseada no princípio de estabelecimento de subconjuntos das variáveis mais representativas de acordo com uma sequência de operações de projeção relacionadas às colunas da matriz de resposta instrumental. O princípio básico do SPA consiste em iniciar com uma variável e, em seguida, incorpora-se outra em cada interação até que um número definido de comprimentos de ondas seja concluído com finalidade de minimizar os problemas de redundância (CHENG *et al.*, 2016).

No SPA, os dados espectrais do conjunto de treinamento são submetidos a uma sequência de operações de projeção que resultam na criação de cadeias de variáveis. Em seguida, as cadeias de variáveis são avaliadas em subconjuntos e passam por um procedimento de eliminação para trás com finalidade de descartar as variáveis não informativas e melhorar a parcimônia do modelo. O mecanismo incorporado ao SPA para minimizar os problemas de colinearidade visa reduzir a propagação do ruído instrumental no processo de classificação para proporcionar modelos estatisticamente estáveis (SOARES *et al.*, 2013).

Os algoritmos de seleção de variáveis por GA e SPA conduzem a seleção dos comprimentos de ondas mais representativos que carregam a informação mais importante relacionada às características em estudo proporcionando um melhor uso da informação espectral e otimizando os resultados de previsão (CHENG *et al.*, 2016). Os procedimentos de seleção de variáveis, por exemplo, GA e SPA, podem ser combinados aos métodos de classificação, como o LDA, e aprimorarem o desempenho do modelo em comparação com o modelo de espectro completo. Esses algoritmos eliminam potenciais interferentes e variáveis que geram uma relação sinal-ruído mais baixa ao selecionar um subconjunto adequado de comprimentos de onda (MARQUES *et al.*, 2015).

O algoritmo genético combinado com a análise discriminante linear (*Genetic Algorithm based Linear Discriminant Analysis – GA-LDA*) seleciona uma série de comprimentos de ondas que descrevem a correlação entre as variáveis preditoras e as variáveis de resposta através de um subconjunto menor de bandas espectrais que têm maior

representatividade de todo o conjunto de dados espectrais para a classificação (MOSCETTI *et al.*, 2014).

Do mesmo modo, o uso do algoritmo de projeções sucessivas em conjunto com a análise discriminante linear (*Successive Projection Algorithm based Linear Discriminant Analysis* – SPA-LDA) objetiva encontrar um pequeno conjunto representativo de variáveis espectrais contendo informações menos redundantes, com ênfase na minimização da colinearidade e do risco médio de erros de classificação (PONTES *et al.*, 2005). LDA é uma ferramenta de classificação que apresenta uma significativa melhoria na capacidade de previsão quando associada a seleção de variáveis gerando modelos de discriminação mais robustos e simples (KHANMOHAMMADI *et al.*, 2013).

Seleção de amostras para o modelo de classificação

A seleção de amostras para os grupos de treinamento e previsão também influencia o desempenho dos modelos construídos e pode ser realizada por rotinas e algoritmos implementados em pacotes computacionais. Vários critérios de seleção são aplicados a diferentes abordagens para seleção de amostras de treinamento representativas, isto é, amostras nos conjuntos de dados uniformemente espalhadas de forma aleatória sobre o intervalo esperado de variabilidade dos dados (WU *et al.*, 1996).

A técnica do algoritmo de Kennard-Stone considera todas as amostras como candidatas para o conjunto de treinamento, escolhidas sequencialmente. Em cada estágio, o objetivo é selecionar as amostras de modo que fiquem uniformemente espaçadas sobre o intervalo de objetos. As duas primeiras amostras são selecionadas escolhendo-se os dois objetos mais distantes. A terceira amostra selecionada é a que está mais distante dos dois primeiros objetos, e assim sucessivamente, empregando as distâncias euclidianas até a última amostra (WU *et al.*, 1996).

O algoritmo de Kennard-Stone tem sido usado em um grande número de situações de planejamento não padrão com capacidade de produzir um delineamento de amostras prático, cujas amostras selecionadas para formar o subconjunto de treinamento representam o máximo da variabilidade presente no conjunto de amostras (KENNARD &

STONE, 1969). Um procedimento de seleção que elege amostras quase igualmente distribuídas sobre o espaço de treinamento levará a uma distribuição plana, tal distribuição é mais favorável do ponto de vista da regressão do que a distribuição normal, de modo que a perda da qualidade preditiva pode ser menor do que o esperado. Portanto, um conjunto de treinamento bem definido é decisivo para determinações químicas precisas em amostras futuras (SHETTY *et al.*, 2012).

Vários estudos utilizando espectroscopia NIR em conjunto às ferramentas de análise multivariada mostram a utilidade dessa técnica para a caracterização de diferentes matérias-primas (RAMBO *et al.*, 2015). O desenvolvimento da tecnologia da computação e da Quimiometria tornaram mais popular e viável a aplicação da espectroscopia de infravermelho próximo atraindo a atenção dos pesquisadores na área de alimentos (CEN & HE, 2007).

4.5. Espectroscopia NIR e Análise Multivariada aplicada à análise de produtos lácteos

Os espectros na região do infravermelho são bastante informativos sobre as matrizes orgânicas, como alimentos. Aplicações desse método na indústria de alimentos datam de 1938, com o trabalho reportando estudos de espectros de absorção para determinação de água em gelatina de Ellis & Bath, atravessando uma lacuna até 1962, quando Hart *et al.* aplicaram a espectroscopia NIR para a determinação de umidade em extratos metanólicos de várias sementes. Desde então, a espectroscopia NIR encontrou um nicho em muitas áreas da indústria de alimentos. O primeiro registro de aplicação de espectroscopia no infravermelho próximo em análises de leite foi de Ben-Gera & Norris descrevendo a análise em leite integral por transmitância (WILLIAMS & STEVENSEN, 1990).

Em leite e derivados, a adição de adulterantes está entre as questões mais problemáticas para qualidade sanitária. Muitos estudos relevantes empregando métodos NIRS foram realizados em leite cru, leite em pó e iogurte com finalidade de detectar adulterações com água e soro de leite (QU *et al.*, 2015). Recentemente, os produtos lácteos foram contaminados com melamina, um produto químico nitrogenado industrial frequentemente usado para fabricação de plástico, papelão, pisos e pratos. Além da

contaminação por melamina, os microrganismos patogênicos também são importantes para o controle de qualidade no leite e a espectroscopia NIR é uma ferramenta alternativa bem sucedida para a detecção rápida de adulterantes e identificação de microrganismos patogênicos em produtos lácteos (KANDPAL *et al.*, 2014; CHEN *et al.*, 2017; ZHANG *et al.*, 2013; AL-QADIRI *et al.*, 2008).

O uso da espectroscopia NIR associada aos métodos quimiométricos é uma ferramenta adicional para o monitoramento de qualidade utilizando análises qualitativas. Liu *et al.* (2011) investigaram a capacidade discriminativa entre leite puro e leite adulterado com ureia, melamina e glicose da espectroscopia NIR combinada a métodos de reconhecimento de padrão empregando o algoritmo SIMCA (*Soft Independent Modeling of Class Analogy*) e PLS-DA para construir os modelos de classificação. O resultado preliminar mostrou que a eficiência da classificação por PLS-DA foi melhor que o método SIMCA. A otimização do modelo de classificação PLS-DA foi realizada pela aplicação do método de suavização com filtro Savitzky-Golay com janela de 5 pontos combinado ao SNV e pela seleção dos comprimentos de onda na região de 1696 a 2500 nm. Os resultados obtidos nessas condições de otimização foi maior que 97% para a classificação entre o leite puro e o leite contendo um dos adulterantes. Para a comparação entre o leite contendo os quatro tipos de adulterantes a taxa de discriminação obtida foi de 93,2%. Dessa forma, a identificação qualitativa para leite adulterado mostrou-se possível, mesmo sem o conhecimento de qual adulterante foi adicionado em consequência do leite adulterado ser facilmente discriminado (LIU *et al.*, 2011).

A espectroscopia de infravermelho próximo (NIR) tem sido aplicada em análises de leite na indústria de laticínios há mais de 30 anos, com aplicações iniciais em produtos em pó e queijos. O desenvolvimento em equipamentos, programas computacionais e técnicas quimiométricas tem aumentado a área de aplicação do método para inclusão de vários tipos de produtos lácteos em toda a cadeia produtiva, ultrapassando os parâmetros de composição tradicionais, como umidade, proteína e gordura, para avaliação de qualidade e rastreamento de origem, por exemplo. As medições de composição e qualidade de queijo pela espectroscopia NIR para avaliação rotineira de parâmetros de composição como gordura, umidade e proteína estão bem estabelecidos, mesmo com as dificuldades apresentadas para análises de queijo baseada na espectroscopia NIR devido a grande variedade de tipos e não homogeneidade da amostra. Em produtos lácteos em pó, o foco da utilização da espectroscopia NIR tem sido a quantificação da composição, explorando a

previsão de proteína e gordura, protocolos de calibração e análises da composição mineral (HOLROYD, 2013).

A medição rápida do leite durante o processamento é um elemento importante da cadeia de processamento de leite. Os artigos que tratam da aplicação de espectroscopia NIR ao leite líquido incluem a determinação da composição do leite, a autenticação dos regimes de alimentação das vacas e a origem geográfica do leite, adicionando informações sobre classificação do leite, conteúdo de microrganismos presente no leite e robustez da calibração. O leite cru é um líquido turvo opaco que apresenta grande dispersão de luz devido a presença de glóbulos de gordura e micelas de caseína em suspensão, o que torna o leite uma matriz desafiadora para estudos empregando o infravermelho próximo. Em consequência, há pouco consenso sobre a gama de comprimentos de onda, escolha entre reflectância ou transmitância e preparação da amostra para análise empregando esse método mesmo existindo um considerável potencial para o aperfeiçoamento da aplicação da espectroscopia NIR aos produtos lácteos (HOLROYD, 2013).

A forte dispersão da luz é um fator complicador da análise em leite cru usando espectros NIR, além das bandas espectrais largas contendo informações sobrepostas e forte absorção de água. A homogeneização dos glóbulos de gordura do leite seria capaz de reduzir e padronizar o espalhamento da radiação para melhorar os resultados da análise multivariada ao produzir uma distribuição de tamanho mais uniforme do glóbulo de gordura, simplificando os efeitos da dispersão sobre os espectros e assim facilitando a análise dos dados. Todavia, técnicas poderosas e avançadas de correção do espalhamento ainda são necessárias para remover a interferência da dispersão nos espectros NIR (BOGOMOLOV & MELENTEVA, 2013; AERNOUTS *et al.*, 2015).

A indústria tem dado bastante importância à necessidade de aprimoramento de sistemas de controle de qualidade não destrutivos e economicamente viáveis. Essa exigência tem aumentado o interesse na espectroscopia NIR utilizando equipamentos portáteis ao oferecer várias vantagens, como portabilidade, desenho ergonômico, simplicidade na análise e robustez, baixo custo e implementação de medições diretas no campo em proporções industriais (SANTOS *et al.*, 2013). Vários estudos com resultados promissores foram realizados usando a espectroscopia NIR portátil em conjunto a PLS-DA para análises em frutas e vegetais principalmente para avaliação do conteúdo de sólidos totais, parâmetros físico-químicos e de qualidade como firmeza (YANG *et al.*, 2012; PÉREZ-MARÍN *et al.*, 2011; PÉREZ-MARÍN *et al.*, 2010). As análises em carnes e

pescados empregando espectroscopia NIR portátil e PLS-DA foram realizadas para classificar os animais baseando-se no regime alimentar e em categorias comerciais, e discriminação entre amostras frescas e congeladas (ZAMORA-ROJAS *et al.*, 2012; FASOLATO *et al.*, 2012). Dessa forma, a espectroscopia NIR portátil mostra-se útil em várias aplicações na indústria de alimentos com resultados próximos aos obtidos pelos métodos convencionais e mais pesquisas aplicando dispositivos NIR portáteis para análises do leite em campo estão em desenvolvimento (SANTOS *et al.*, 2013).

Observa-se que a espectroscopia NIR pode ser utilizada para diferenciação entre variedades de matérias-primas e classificação de acordo com diferentes origens ou técnicas de processamento, entre outros parâmetros, quando combinada a análise multivariada de dados apropriada, exemplificando a versatilidade e relevância da técnica, sendo necessário o desenvolvimento de modelos de classificação robustos (QU *et al.*, 2015).

A literatura científica apresenta-se limitada sobre a aplicação da classificação multivariada associada à espectroscopia NIR em amostras de leite UHT, apresentando vários trabalhos empregando NIR na quantificação dos principais componentes do leite com calibração multivariada. Apesar dos trabalhos relatados na literatura apresentarem informação diversa e não consensual quanto ao modo de medida, caminho ótico, número de varreduras e faixa de comprimento de onda empregados na análise de componentes como a lactose no leite, esses estudos demonstram a capacidade da espectroscopia NIR ser utilizada com acurácia para a classificação de amostras complexas.

A aplicabilidade para análises rápidas da composição do leite UHT usando a espectroscopia NIR e classificação multivariada para avaliar a qualidade do leite em relação a presença de lactose expressa viabilidade, pois um modelo de classificação satisfatório seria capaz de reconhecer a presença ou ausência desse carboidrato. Dessa forma, considerando os parâmetros mais adequados para aquisição dos espectros e a construção de modelos de classificação bem sucedidos, a espectroscopia NIR empregando equipamentos de bancada e portátil em conjunto a análise multivariada pode ser uma ferramenta capaz de classificar corretamente os leites com e sem lactose.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- 1) AERNOOTS, B.; VAN BEERS, R.; WATTÉ, R.; HUYBRECHTS, T.; LAMMERTYN, J.; SAEYS, W. Visible and near-infrared bulk optical properties of raw milk. **Journal of Dairy Science**. v.98, n.10, p.6727–6738, out. 2015.
- 2) AL-QADIRI, H. M.; LIN, M.; AL-HOLY, M. A.; CAVINATO, A. G.; RASCO, B. A. Monitoring Quality Loss of Pasteurized Skim Milk Using Visible and Short Wavelength Near-Infrared Spectroscopy and Multivariate Analysis. **Journal of Dairy Science**. v.91, n.3, p.950–958, mar. 2008.
- 3) AMADOR-ESPEJO, G. G.; GALLARDO-CHACON, J. J.; NYKÄNEN, H.; JUAN, B.; TRUJILLO, A. J. Effect of ultra high-pressure homogenization on hydro- and liposoluble milk vitamins. **Food Research International**. v.77, n.1, p.49–54, nov. 2015.
- 4) ANSARI, S. A.; SATAR, R. Recombinant β -galactosidases – Past, present and future: A mini review. **Journal of Molecular Catalysis B: Enzymatic**. v.81, p.1–6, set. 2012.
- 5) AOAC. Official Methods of Analysis of the Association of Official Analytical Chemists. AOAC Official Method 896.01-1896. Lactose in milk. Polarimetric Method. **AOAC**, 2005.
- 6) AOAC. Official Methods of Analysis of the Association of Official Analytical Chemists. AOAC Official Method 972.16-1972. Fat, lactose, protein, and solids in milk. Mid-Infrared Spectroscopic Method. **AOAC**, 2005.
- 7) AYVAZ, H.; RODRIGUEZ-SAONA, L. E. Application of handheld and portable spectrometers for screening acrylamide content in commercial potato chips. **Food Chemistry**. v.174, p.154–162, mai. 2015.
- 8) BAHNA, S. L. Is it milk allergy or lactose intolerance? **Immunology and Allergy Clinics of North America**. v.16, n.1, p.187-198, fev. 1996.
- 9) BALABIN, R. M.; SMIRNOV, S. V. Variable selection in near-infrared spectroscopy: Benchmarking of feature selection methods on biodiesel data. **Analytica Chimica Acta**. v.692, n.1–2, p.63–72, abr. 2011.
- 10) BASRI, K. N.; HUSSAIN, M. N.; BAKAR, J.; SHARIF, Z.; KHIR, M. F. A.; ZOOLFAKAR, A. S. Classification and quantification of palm oil adulteration via

- portable NIR spectroscopy. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**. v.173, p.335–342, fev. 2017.
- 11) BERRUETA, L. A.; ALONSO-SALCES, R. M.; HÉBERGER, K. Supervised pattern recognition in food analysis. **Journal of Chromatography A**. v.158, n.1–2, p.196–214, jul. 2007.
- 12) BOGOMOLOV, A.; MELENTEVA, A. Scatter-based quantitative spectroscopic analysis of milk fat and total protein in the region 400–1100 nm in the presence of fat globule size variability. **Chemometrics and Intelligent Laboratory Systems**. v.126, p.129–139, jul. 2013.
- 13) BORRÀS, E.; AMIGO, J. M.; BERG, F. V. D.; BOQUÉ, R.; BUSTO, O. Fast and robust discrimination of almonds (*Prunus amygdalus*) with respect to their bitterness by using near infrared and partial least squares-discriminant analysis. **Food Chemistry**. v.153, p.15–19, jun. 2014.
- 14) BRASIL. Decreto nº 30.691, de 29 de março de 1952: Regulamento da inspeção industrial e sanitária de produtos de origem animal – RIISPOA. TÍTULO VIII - Inspeção industrial e sanitária do leite e derivados. Ministério da Agricultura, Pecuária e Abastecimento, Brasil. **Diário Oficial da União**, 07/07/1952.
- 15) BRASIL. Instrução Normativa nº 62, de 29 de dezembro de 2011: Anexo II – “Anexo IV - Regulamento técnico de identidade e qualidade de leite cru refrigerado”. Ministério da Agricultura, Pecuária e Abastecimento, Brasil. **Diário Oficial da União**, 30/12/2011.
- 16) BRASIL. Lei nº 13.305, de 4 de julho de 2016: Acrescenta art. 19-A ao Decreto-Lei nº986, de 21 de outubro de 1969, que “institui normas básicas sobre alimentos”, para dispor sobre a rotulagem de alimentos que contenham lactose. Casa Civil, Brasil. **Diário Oficial da União**, 05/07/2016.
- 17) BRASIL. Resolução RDC nº 135, de 8 de fevereiro de 2017: Altera a Portaria SVS/MS nº 29, de 13 de janeiro de 1998, que Aprova o Regulamento Técnico Referente a Alimentos para Fins Especiais, para Dispor Sobre os Alimentos para Dietas com Restrição de Lactose. Agência Nacional de Vigilância Sanitária, Brasil. **Diário Oficial da União**, 09/02/2017.
- 18) BRASIL. Portaria nº 146, de 07 de março de 1996: Anexo XII - Regulamento técnico de identidade e qualidade do leite UAT (UHT). Ministério da Agricultura, Pecuária e Abastecimento, Brasil. **Diário Oficial da União**, 11/03/1996.

- 19) BRITO, A. L. B.; BRITO, L. R.; HONORATO, F. A.; PONTES, M. J. C.; PONTES, L. F. B. L. Classification of cereal bars using near infrared spectroscopy and linear discriminant analysis. **Food Research International**. v.51, n.2, p.924–928, mai. 2013.
- 20) BROWN-ESTERS, O.; MC NAMARA, P.; SAVAIANO, D. Dietary and biological factors influencing lactose intolerance. **International Dairy Journal**. v.22, n.2, p.98–103, fev. 2012.
- 21) BUNACIU, A. A.; ABOUL-ENEIN, H. Y.; HOANG, V. D. Vibrational spectroscopy used in milk products analysis: A review. **Food Chemistry**. v.196, p.877–884, abr. 2016.
- 22) BOYSWORTH, M. K.; BOOKSH, K. S. Aspects of multivariate calibration applied to near-infrared spectroscopy. In: BURNS, D. A.; CIURCZAK, E. W. Handbook of near-infrared analysis / edited by Donald A. Burns and Emil W. Ciurczak. **CRC press**. Taylor & Francis Group. 3rd ed. 834p., p.207–229, 2007.
- 23) CAYUELA, J. A.; WEILAND, C. Intact orange quality prediction with two portable NIR spectrometers. **Postharvest Biology and Technology**. v.58, n.2, p.113–120, nov. 2010.
- 24) CEN, H.; HE, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. **Trends in Food Science & Technology**. v.18, n.2, p.72–83, fev. 2007.
- 25) CHEN, H.; TAN, C.; LIN, Z.; WU, T. Detection of melamine adulteration in milk by near-infrared spectroscopy and one-class partial least squares. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**. v.173, p.832–836, fev. 2017.
- 26) CHEN, H.; PAN, T.; CHEN, J.; LU, Q. Waveband selection for NIR spectroscopy analysis of soil organic matter based on SG smoothing and MWPLS methods. **Chemometrics and Intelligent Laboratory Systems**. v.107, n.1, p.139–146, mai. 2011.
- 27) CHEN, W.; CHEN, H.; XIA, Y.; YANG, J.; ZHAO, J.; TIAN, F.; ZHANG, H. P.; ZHANG, H. Immobilization of recombinant thermostable beta-galactosidase from *Bacillus stearothermophilus* for lactose hydrolysis in milk. **Journal of Dairy Science**. v.92, n.2, p.491–498, fev. 2009.
- 28) CHENG, J.-H.; SUN, D.-W.; PU, H. Combining the genetic algorithm and successive projection algorithm for the selection of feature wavelengths to evaluate exudative characteristics in frozen–thawed fish muscle. **Food Chemistry**. v.197, p.855–863, abr. 2016.

- 29) COLLELL, C.; GOU, P.; ARNAU, J.; MUÑOZ, I.; COMAPOSADA, J. NIR technology for on-line determination of superficial a_w and moisture content during the drying process of fermented sausages. **Food Chemistry**. v.135, n.3, p.1750–1755, dez. 2012.
- 30) COLUMBIA UNIVERSITY [online]. Intensive Seminars in Moderns Chemistry. Experiment 7 – IR Spectroscopy. **Columbia University**, New York - NY. 2007. Disponível em: http://www.columbia.edu/cu/chemistry/ugrad/hssp/EXP_7.html. Acesso em: 25 de janeiro de 2017.
- 31) COZZOLINO, D. The role of vibrational spectroscopy as a tool to assess economically motivated fraud and counterfeit issues in agricultural products and foods. **Anal. Methods**. v.7, n.22, p.9390–9400, out. 2015.
- 32) COZZOLINO, D.; CYNKAR, W. U.; SHAH, N.; SMITH, P. Multivariate data analysis applied to spectroscopy: potential application to juice and fruit quality. **Food Research International**. v.44, n.7, p.1888–1896, ago. 2011.
- 33) DENG, Y.; MISSELWITZ, B.; DAI, N.; FOX, M. Lactose intolerance in adults: biological mechanism and dietary management. **Nutrients**. v.7, n.9, p.8020-8035, set. 2015.
- 34) DUBOIS, M.; GILLES, K. A.; HAMILTON, J. K.; REBERS, P. A.; SMITH, F. Colorimetric Method for Determination of Sugars and Related Substances. **Anal. Chem.** v.28, n.3, p.350–356, mar. 1956.
- 35) ELLIOTT, J. A.; EMMONS, D. B.; YATES, A. R. The Influence of the Bacterial Quality of Milk on the Properties of Dairy Products. A Review. **Canadian Institute of Food Science and Technology Journal**. v.7, n.1, p.32-39, jan. 1974.
- 36) ENGEL, J.; GERRETZEN, J.; SZYMAŃSKA, E.; JANSEN, J. J.; DOWNEY, G.; BLANCHET, L.; BUYDENS, L. M. C. Breaking with trends in pre-processing? **TrAC - Trends in Analytical Chemistry**. v.50, p.96–106, out. 2013.
- 37) EUROPEAN COMMUNITY (EC). COUNCIL DIRECTIVE 92/46/EEC of 16 June 1992. Laying down the health rules for the production and placing on the market of raw milk, heat-treated milk and milk-based products. **Official Journal of the European Communities**, 14/09/1992.
- 38) EUROPEAN COMMUNITY (EC). COUNCIL REGULATION (EC) No 2597 /97 of 18 December 1997. Laying down additional rules on the common organization of the market in milk and milk products for drinking milk. **Official Journal of the European Communities**, 23/12/1997.

- 39) FARINA, E. M. M. Q.; GUTMAN, G. E.; LAVARELLO, P. J.; NUNES, R.; REARDON, T. Private and public milk standards in Argentina and Brazil. **Food Policy**. v.30, n.3, p.302–315, jun. 2005.
- 40) FASOLATO, L.; BALZAN, S.; RIOVANTO, R.; BERZAGHI, P.; MIRISOLA, M.; FERLITO, J. C.; SERVA, L.; BENOZZO, F.; PASSERA, R.; TEPEDINO, V.; NOVELLI, E. Comparison of Visible and Near-Infrared Reflectance Spectroscopy to Authenticate Fresh and Frozen-Thawed Swordfish (*Xiphias gladius* L). **Journal of Aquatic Food Product Technology**. v.21, n.5, p.493-507, jan. 2012.
- 41) FERREIRA, M. M. C. Preparação dos dados para análise. In: Ferreira, M. M. C. Quimiometria – conceitos, métodos e aplicações / Márcia Miguel Castro Ferreira – Campinas, SP: **Editora da Unicamp**. 496p., p.29-105, 2015. (a)
- 42) FERREIRA, M. M. C. Análise exploratória de dados. In: Ferreira, M. M. C. Quimiometria – conceitos, métodos e aplicações / Márcia Miguel Castro Ferreira – Campinas, SP: **Editora da Unicamp**. 496p., p.107-250, 2015. (b)
- 43) GALLO, M.; FERRANTI, P. The evolution of analytical chemistry methods in foodomics. **Journal of Chromatography A**. v.1428, p.3–15, jan. 2016.
- 44) GÄNZLE, M. G.; HAASE, G.; JELEN, P. Lactose: Crystallization, hydrolysis and value-added derivatives. **International Dairy Journal**. v.18, n.7, p.685–694, jun. 2008.
- 45) GASPARDO, B.; ZOTTO, S. DEL; TORELLI, E.; CIVIDINO, S. R.; FIRRAO, G.; RICCIA, G. DELLA; STEFANON, B. A rapid method for detection of fumonisins B₁ and B₂ in corn meal using Fourier transform near infrared (FT-NIR) spectroscopy implemented with integrating sphere. **Food Chemistry**. v.135, n.3, p.1608–1612, dez. 2012.
- 46) GORRY, P. A. General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method. **Anal. Chem.** v.62, n.6, p.570–573, mar. 1990.
- 47) HAMMOND, L. D. Determination of lactose alone and in the presence of sucrose by the method of Munson and Walker. **Journal of Research of the National Bureau of Standards**. v.41, p.211–221, set. 1948.
- 48) HARJU, M.; KALLIOINEN, H.; TOSSAVAINEN, O. Lactose hydrolysis and other conversions in dairy products: Technological aspects. **International Dairy Journal**. v.22, n.2, p.104-109, fev. 2012.
- 49) HAUG, A.; HØSTMARK, A. T.; HARSTAD, O. M. Bovine milk in human nutrition – a review. **Lipids in Health and Disease**. v.6, n.25, set. 2007.

- 50) HUANG, Y.; WU, Z.; SU, R.; RUAN, G.; DU, F.; LI, G. Current application of chemometrics in traditional Chinese herbal medicine research. **Journal of Chromatography B**. v.1026, p.27–35, jul. 2016.
- 51) HUCK, C. W. Advances of infrared spectroscopy in natural product research. **Phytochemistry Letters** v.11, p.384–393, mar. 2015.
- 52) HUPPERTZ, T.; GAZI, I. Lactose in dairy ingredients: Effect on processing and storage stability. **Journal of Dairy Science**. v.99, n.8, p.6842–6851, ago. 2016.
- 53) HUSAIN, Q. β -Galactosidases and their potential applications: a review. **Critical Reviews in Biotechnology**. v.30, n.1, p.41–62, mar. 2010.
- 54) ILLANES, A. Lactose: Production and Upgrading. In: Illanes, A.; Guerrero, C.; Vera, C.; Wilson, L.; Conejeros, R.; Scott, F. Lactose-Derived Prebiotics - A Process Perspective. **Academic Press**. 312p., p.1-33, 2016.
- 55) IWASAKI, K.; NAKAJIMA, M.; NAKAO, S. Galacto-oligosaccharide production from lactose by an enzymic batch reaction using β -galactosidase. **Process Biochemistry**. v.31, n.1, p.69–76, 1996.
- 56) JELEN, P.; TOSSAVAINEN, O. Low lactose and lactose-free milk and dairy products – prospects, technologies and applications. **Australian Journal of Dairy Technology**. v.58, n.2, p.161–165, ago. 2003.
- 57) KANDPAL, L. M.; CHO, B.-K. A Review of the Applications of Spectroscopy for the Detection of Microbial Contaminations and Defects in Agro Foods. **Journal of Biosystems Engineering**. v.39, n.3, p.215–226, ago. 2014.
- 58) KAROUI, R.; BAERDEMAEKER, J. DE. A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products. **Food Chemistry**. v.102, n.3, p.621-640, 2007.
- 59) KENNARD, R. W.; STONE, L. A. Computer aided design of experiments. **Technometrics**. v.11, n.1, p.137–148, 1969.
- 60) KHANMOHAMMADI, M.; GARMARUDI, A. B.; GUARDIA, M. DE LA. Feature selection strategies for quality screening of diesel samples by infrared spectrometry and linear discriminant analysis. **Talanta**. v.104, p.128–134, jan. 2013.
- 61) KUMAR, N.; BANSAL, A.; SARMA, G. S.; RAWAL, R. K. Chemometrics tools used in analytical chemistry: An overview. **Talanta**. v.123, p.186–199, jun. 2014.
- 62) LANE, J. H.; EYNON, L. Determination of reducing sugars by means of Fehling's solution with methylene blue as internal indicator. **J. Soc. Chem. Ind.** v.42, p.32–36, 1923.

- 63) LIU, H.; ZHAO, K.; LIU, J. Effects of Glucose Availability on Expression of the Key Genes Involved in Synthesis of Milk Fat, Lactose and Glucose Metabolism in Bovine Mammary Epithelial Cells. **PLoS ONE**. v.8, n.6, p.1-6, jun. 2013.
- 64) LIU, R.; LV, G.; HE, B.; XU, K. Discriminant analysis of milk adulteration based on near-infrared spectroscopy and pattern recognition. **Proceedings of SPIE - The International Society for Optical Engineering**. v.7906, fev. 2011. doi:10.1117/12.873260.
- 65) LULE, V. K.; GARG, S.; TOMAR, S. K. Food intolerance: lactose intolerance. In: CABALLERO, B.; FINGLAS, P.M.; TOLDRÁ, F. (eds.) Encyclopedia of Food and Health. **Academic Press**. p.43-48, 2016.
- 66) MARQUES, A. S.; MORAES, E. P.; JÚNIOR, M. A. A.; MOURA, A. D.; NETO, V. F. A.; NETO, R. M.; LIMA, K. M. G. Rapid discrimination of *klebsiella pneumoniae* carbapenemase 2 – producing and non-producing *klebsiella pneumoniae* strains using near-infrared spectroscopy (NIRS) and multivariate analysis. **Talanta**. v.134, p.126–131, mar. 2015.
- 67) MARQUETTI, I.; LINK, J. V.; LEMES, A. L. G.; SCHOLZ, M. B. S.; VALDERRAMA, P.; BONA, E. Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee. **Computers and Eletronics in Agriculture**. v.121, p.313–319, fev. 2016.
- 68) MATHER, I. H. Mammary Gland, Milk Biosynthesis and Secretion: Secretion of Milk Constituents. **Encyclopedia of Dairy Sciences (Second Edition)**. p.373–380, dez. 2011.
- 69) MCGORRIN, R. J. One Hundred Years of Progress in Food Analysis. **J. Agric. Food Chem**. v.57, p.8076–8088, ago. 2009.
- 70) MESSIA, M.C.; CANDIGLIOTA, T.; MARCONI, E. Assessment of quality and technological characterization of lactose-hydrolyzed milk. **Food Chemistry**. v.104, n.3, p.910–917, 2007.
- 71) MICHALSKI, M. C.; JANUEL, C. Does homogenization affect the human health properties of cow's milk? **Trends in Food Science & Technology**. v.17, n.8, p.423-437, ago. 2006.
- 72) MICIŃSKI, J.; KOWALSKI, I. M.; ZWIERZCHOWSKI, G.; SZAREK, J.; PIEROŻYŃSKI, B.; ZABŁOCKA, E. Characteristics of cow's milk proteins including

- allergenic properties and methods for its reduction. **Polish Annals of Medicine**. v.20, n.1, p.69–76, set. 2013.
- 73) MILLER, G. L. Use of dinitrosalicylic acid reagent for determination of reducing sugar. **Analytical Chemistry**. v.31, n.3, p.426–428, mar. 1959.
- 74) MISSELWITZ, B.; POHL, D.; FRÜHAUF, H.; FRIED, M.; VAVRICKA, S. R.; FOX, M. Lactose malabsorption and intolerance: pathogenesis, diagnosis and treatment. **United European Gastroenterology Journal**. v.1, n.3, p.151-159, mar. 2013.
- 75) MODROÑO, S.; SOLDADO, A.; MARTÍNEZ-FERNÁNDEZ, A.; ROZADDELGADO, B. DE LA. Handheld NIRS sensors for routine compound feed quality control: Real time analysis and field monitoring. **Talanta**. v.162, p.597–603, jan. 2017.
- 76) MOREIRA, E. D. T.; PONTES, M. J. C.; GALVÃO, R. K. H.; ARAÚJO, M. C. U. Near infrared reflectance spectrometry classification of cigarettes using the successive projections algorithm for variable selection. **Talanta**. v.79, n.5, p.1260–1264, out. 2009.
- 77) MORLOCK, G. E.; MORLOCK, L. P.; LEMO, C. Streamlined analysis of lactose-free dairy products. **Journal of Chromatography A**. v.1324, p.215-223, jan. 2014.
- 78) MOSCETTI, R.; HAFF, R. P.; SRANWONG, S.; MONARCA, D.; CECCHINI, M.; MASSANTINI, R. Nondestructive detection of insect infested chestnuts based on NIR spectroscopy. **Postharvest Biology and Technology**. v.87, p.88–94, jan. 2014.
- 79) NELSON, N. A photometric adaptation of the Somogyi method for the determination of glucose. **J. Biol. Chem.** v.153, p.375–80, fev. 1944.
- 80) NETCHACOVITCH, L.; THIRY, J.; DE BLEYE, C.; CHAVEZ, P.-F.; KRIER, F.; SACRÉ, P.-Y.; EVRARD, B.; HUBERT, PH.; ZIEMONS, E. Vibrational spectroscopy and microspectroscopy analyzing qualitatively and quantitatively pharmaceutical hot melt extrudates. **Journal of Pharmaceutical and Biomedical Analysis**. v.113, p.21–33, set. 2015.
- 81) NIE, Y.; HUANG, M.; LU, J.; QIAN, J.; LIN, W.; CHU, J.; ZHUANG, Y.; ZHANG, S. Impacts of high β -galactosidase expression on central metabolism of recombinant *Pichia pastoris* GS115 using glucose as sole carbon source via ^{13}C metabolic flux analysis. **Journal of Biotechnology**. v.187, p.124–134, out. 2014.
- 82) NÚÑEZ-SÁNCHEZ, N.; MARTÍNEZ-MARÍN, A. L.; POLVILLO, O.; FERNÁNDEZ-CABANÁS, V. M.; CARRIZOSA, J.; URRUTIA, B.; SERRADILLA,

- J. M. Near Infrared Spectroscopy (NIRS) for the determination of the milk fat fatty acid profile of goats. **Food Chemistry**. v.190, p.244–252, jan. 2016.
- 83) OCHI, H.; SUMI, M.; NAKATA, I.; SAITO, H.; UOZUMI, M.; IWATSUKI, K. Sensometric calibration of sensory characteristics of commercially available milk products with instrumental data. **J. Dairy Sci.** v.93, n.5, p.1794–1806, mai. 2010.
- 84) PASQUINI, C. Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. **J. Braz. Chem. Soc.** v.14, n.2, p.198–219, 2003.
- 85) PEREIRA, P. C. Milk nutritional composition and its role in human health. **Nutrition**. v.30, n.6, p.619–627, jun. 2014.
- 86) PÉREZ-MARIN, D.; PAZ, P.; GUERRERO, J.-E.; GARRIDO-VARO, A.; SÁNCHEZ, M.-T. Miniature handheld NIR sensor for the on-site non-destructive assessment of post-harvest quality and refrigerated storage behavior in plums. **Journal of Food Engineering**. v.99, n.3, p.294–302, ago. 2010.
- 87) PÉREZ-MARIN, D.; SÁNCHEZ, M.-T.; PAZ, P.; GONZÁLEZ-DUGO, V.; SORIANO, M.-A. Postharvest shelf-life discrimination of nectarines produced under different irrigation strategies using NIR-spectroscopy. **LWT - Food Science and Technology**. v.44, n.6, p.1405-1414, jul. 2011.
- 88) PLANS, M.; SIMÓ, J.; CASAÑAS, F.; SABATÉ, J.; RODRIGUEZ-SAONA, L. Characterization of common beans (*Phaseolus vulgaris* L.) by infrared spectroscopy: Comparison of MIR, FT-NIR and dispersive NIR using portable and benchtop instruments. **Food Research International**. v.54, n.2, p.1643–1651, dez. 2013.
- 89) PONTES, M. J. C.; GALVÃO, R. K. H.; ARAÚJO, M. C. U.; MOREIRA, P. N. T.; NETO, O. D. P.; JOSÉ, G. E.; SALDANHA, T. C. B. The successive projections algorithm for spectral variable selection in classification problems. **Chemometrics and Intelligent Laboratory Systems**. v.78, n.1–2, p.11–18, jul. 2005.
- 90) POREP, J. U.; KAMMERER, D. R.; CARLE, R. On-line application of near infrared (NIR) spectroscopy in food production. **Trends in Food Science & Technology**. v.46, n.2, p.211-230, dez. 2015.
- 91) POULSEN, N. A.; RYBICKA, I.; POULSEN, H. D.; LARSEN, L. B.; ANDERSEN, K. K.; LARSEN, M. K. Seasonal variation in content of riboflavin and major minerals in bulk milk from three Danish dairies. **International Dairy Journal**. v.42, p.6–11, mar. 2015.

- 92) QI, P. X.; REN, D.; XIAO, Y.; TOMASULA, P. M. Effect of homogenization and pasteurization on the structure and stability of whey protein in milk. **Journal of Dairy Science**. v.98, n.5, p.2884-2897, mai. 2015.
- 93) QU, J.-H.; LIU, D.; CHENG, J.-H.; SUN, D.-W.; MA, J.; PU, H.; ZENG, X.-A. Applications of Near-infrared Spectroscopy in Food Safety Evaluation and Control: A Review of Recent Research Advances. **Critical Reviews in Food Science and Nutrition**. v.55, n.13, p.1939–1954, 2015.
- 94) QUIGLEY, L.; O'SULLIVAN, O.; BERESFORD, T. P.; PAUL ROSS, R.; FITZGERALD, G. F.; COTTER, P. D. Molecular approaches to analysing the microbial composition of raw milk and raw milk cheese. **International Journal of Food Microbiology**. v.150, n.2–3, p.81–94, nov. 2011.
- 95) RAIKOS, V. Effect of heat treatment on milk protein functionality at emulsion interfaces. A review. **Food Hydrocolloids**. v.24, n.4, p.259-265, jun. 2010.
- 96) RAJALAHTI, T.; KVALHEIM, O. M. Multivariate data analysis in pharmaceuticals: a tutorial review. **International Journal of Pharmaceutics**. v.417, n.1–2, p.280–290, set. 2011.
- 97) RAMBO, M. K. D.; ALVES, A. R.; GARCIA, W. T.; FERREIRA, M. M. C. Multivariate analysis of coconut residues by near infrared spectroscopy. **Talanta**. v.138, p.263–272, jun. 2015.
- 98) REID, L. M.; O'DONNELL, C. P.; DOWNEY G. Recent technological advances for the determination of food authenticity. **Trends in Food Science & Technology**. v.17, n.7, p.344–353, jul. 2006.
- 99) REINHOLDS, I.; BARTKEVICS, V.; SILVIS, I. C. J.; RUTH, S. M. VAN; ESSLINGER, S. Analytical techniques combined with chemometrics for authentication and determination of contaminants in condiments: A review. **Journal of Food Composition and Analysis**. v.44, p.56–72, dez. 2015.
- 100) RINNAN, A.; BERG, F. VAN DEN; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **TrAC - Trends in Analytical Chemistry**. v.28, n.10, p.1201-1222, nov. 2009.
- 101) ROGGO, Y.; CHALUS, P.; MAURER, L.; LEMA-MARTINEZ, C.; EDMOND, A.; JENT N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. **Journal of Pharmaceutical and Biomedical Analysis**. v.44, n.3, p.683–700, jul. 2007.

- 102) SANTOS, C. A. T. DOS; LOPO, M.; PÁSCOA, R. N. M. J; LOPES, J. A. A review on the applications of portable near-infrared spectrometers in the agro-food industry. **Applied Spectroscopy**. v.67, n.11, p.1215–1233, nov. 2013.
- 103) SAKKAS, L.; MOUTAFI, A.; MOSCHOPOULOU, E.; MOATSOU, G. Assessment of heat treatment of various types of milk. **Food Chemistry**. v.159, p.293-301, set. 2014.
- 104) SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**. v.36, n.8, p.1627-1639, jul. 1964.
- 105) SCHOORL, N. Suiker-titratie. **Chemisch Weekblad**. p.130–134, 1929.
- 106) SCOTTER, C. Use of near infrared spectroscopy in the food industry with particular reference to its applications to on/in-line food processes. **Food Control**. v.1, n.3, p.142-149, jul. 1990.
- 107) SHENDURSE, A. M.; KHEDKAR, C. D. Lactose. In: CABALLERO, B., FINGLAS, P., AND TOLDRÁ, F. (eds.) The Encyclopedia of Food and Health. Oxford: **Academic Press**. v.3, p.509-516, dez. 2016.
- 108) SHETTY, N.; RINNAN, A.; GISLUM, R. Selection of representative calibration sample sets for near-infrared reflectance spectroscopy to predict nitrogen concentration in grasses. **Chemometrics and Intelligent Laboratory Systems**. v.111, n.1, p.59–65, fev. 2012.
- 109) SILVA, C. S.; BORBA, F. S. L.; PIMENTEL, M. F.; PONTES, M. J. C.; HONORATO, R. S.; PASQUINI, C. Classification of blue pen ink using infrared spectroscopy and linear discriminant analysis. **Microchemical Journal**. v.109, p.122–127, jul. 2013.
- 110) SOARES, S. F. C.; GOME, A. A.; FILHO, A. R. G.; ARAUJO, M. C. U.; GALVÃO, R. K. H. The successive projections algorithm. **Trends in Analytical Chemistry**. v.42, p.84-98, jan. 2013.
- 111) SOLA-LARRAÑAGA, C.; NAVARRO-BLASCO, I. Chemometric analysis of minerals and trace elements in raw cow milk from the community of Navarra, Spain. **Food Chemistry**. v.112, n.1, p.189–196, jan. 2009.
- 112) SRIVASTAVA, A.; TRIPATHI, R.; VERMA, S.; SRIVASTAVA, N.; RAWAT, A. K. S.; DEEPAK, D. A novel method for quantification of lactose in mammalian milk through HPTLC and determination by a mass spectrometric technique. **Anal. Methods**. v.6, p.7268–7276, jul. 2014.

- 113) STANIMIROVA, I.; ÜSTÜN, B.; CAJKA, T.; RIDDELLOVA, K.; HAJŠLOVA, J.; BUYDENS, L. M. C.; WALCZAK, B. Tracing the geographical origin of honeys based on volatile compounds profiles assessment using pattern recognition techniques. **Food Chemistry**. v.118, n.1, p.171–176, jan. 2010.
- 114) THIEBAUD, M.; DUMAY, E.; PICART, L.; GUIRAUD, J. P.; CHEFTEL, J. C. High-pressure homogenization of raw bovine milk. Effects on fat globule size distribution and microbial inactivation. **International Dairy Journal**. v.13, n.6, p.427–439, 2003.
- 115) THOMAS, E. L. Trends in milk flavors. **Journal of Dairy Science**. v.64, n.6, p.1023–1027, 1981.
- 116) TROISE, A. D.; BANDINI, E.; DONNO, R. DE; MEIJER, G.; TREZZI, M.; FOGLIANO, V. The quality of low lactose milk is affected by the side proteolytic activity of the lactase used in the production process. **Food Research International**. v.89, n.1, p.514–525, nov. 2016.
- 117) TUNICK, M. H.; REN, D. X.; VAN HEKKEN, D. L.; BONNAILLIE, L.; PAUL, M.; KWOCZAK, R.; TOMASULA, P. M. Effect of heat and homogenization on in vitro digestion milk. **Journal of Dairy Science**. v.99, n.6, p.4124–4139, jun. 2016.
- 118) WALSTRA, P.; WOUTERS, J. T. M.; GEURTS, T. J. Milk: Main Characteristics. In: WALSTRA, P.; WOUTERS, J. T. M.; GEURTS, T. J. Dairy science and technology. 2nd ed. **CRC Press**. 808p, p.3–16, 2005.
- 119) WILLIAMS, P. C.; STEVENSEN, S. G. Near-infrared reflectance analysis: food industry applications. **Trends in Food Science & Technology**. v.1, p.44–48, jul. 1990.
- 120) WILSON, J. Milk intolerance: lactose intolerance and cow's milk protein allergy. **Newborn and Infant Nursing Reviews**. v.5, n.4, p.203–207, dez. 2005.
- 121) WORKMAN JR., J.; WEYER, L. Introduction to near-infrared spectra. In: WORKMAN JR., J.; WEYER, L. Practical guide and spectral atlas for interpretive near-infrared spectroscopy. **CRC Press**. 2nd edition. 326p., p.1–17, 2012.
- 122) WU, W.; WALCZAK, B.; MASSART, D. L.; HEUERDING, S.; ERNI, F.; LAST, I. R.; PREBBLE, K. A. Artificial neural networks in classification of NIR spectral data: Design of the training set. **Chemometrics and Intelligent Laboratory Systems**. v.33, n.1, p.35–46, mai. 1996.

- 123) XIAOBO, Z.; JIEWEN, Z.; POVEY, M. J. W.; HOLMES, M.; HAMPIN, M. Variables selection methods in near-infrared spectroscopy. **Analytica Chimica Acta**. v.667, n.1–2, p.14–32, mai. 2010.
- 124) YANG, F.; QIU, X.-Z.; HAO, R.; GAO, F.; DU, W.; ZHANG, Z.-Y. Determination of Acidity in Oranges Based on Emphatic Orthogonal Signal Correction and Principal Component Orthogonal Signal Correction. **Spectroscopy and Spectral Analysis**. v.32, n.7, p.1931-1934, jul. 2012.
- 125) YAYOTA, M.; TSUKAMOTO, M.; YAMADA, Y.; OHTAN, S. Milk composition and flavor under different feeding systems: A survey of dairy farms. **J. Dairy Sci**. v.96, n.8, p.5174–5183, 2013.
- 126) YEOW, Y. L.; AZALI, S.; OW, S.Y.; WONG, M. C. L.; LEONG, Y.-K. Evaluating the third and fourth derivatives of spectral data. **Talanta**. v.68, n.2, p.156-164, dez. 2005.
- 127) ZAMORA-ROJAS, E.; PÉREZ-MARÍN, D.; PEDRO-SANZ, E. DE; GUERRERO-GINEL, J. E.; GARRIDO-VARO, A. In-situ iberian pig carcass classification using a micro-electro-mechanical system (MEMS)-based near infrared (NIR) spectrometer. **Meat Science**. v.90, n.3, p.636-642, mar. 2012.
- 128) ZEAITER, M.; ROGER, J.-M.; BELLON-MAUREL, V. Robustness of models developed by multivariate calibration. Part II: The influence of pre-processing methods. **TrAC - Trends in Analytical Chemistry**. v.24, n.5, p.437-445, mai. 2005.
- 129) ZHANG, L.-G.; ZHANG, X.; NI, L.-J.; XUE, Z.-B.; GU, X.; HUANG, S.-H. Rapid identification of adulterated cow milk by non-linear pattern recognition methods based on near infrared spectroscopy. **Food Chemistry**. v.145, p.342–348, ago. 2013.
- 130) ZHOU, Q. Z.; CHEN, X. D.; LI, X. Kinetics of Lactose Hydrolysis by β -Galactosidase of *Kluyveromyces lactis* Immobilized on Cotton Fabric. **Biotechnology and Bioengineering**. v.81, n.2, p.127–133, jan. 2003.

6. ARTIGO ORIGINAL

Classificação Multivariada de leite UHT quanto à presença de lactose utilizando espectrômetros NIR portátil e de bancada

Resumo: O leite com lactose hidrolisada foi desenvolvido a partir dos anos 1970 a fim de atender aos indivíduos com intolerância a este carboidrato. Assim, a demanda por produtos sem lactose pelos consumidores intolerantes fez surgir um segmento de mercado para esse gênero alimentício, cujo controle de qualidade tem que ser garantido. A fim de avaliar amostras de leite quanto ao conteúdo de lactose, este trabalho propõe uma metodologia analítica para classificar leites UHT comum e sem lactose utilizando espectros NIR obtidos em espectrômetros FT-NIR de bancada e NIR ultracompacto, útil para análise em campo, e métodos de classificação multivariada. Para isso, um total de 71 amostras foi adquirido, sendo 41 de leite UHT sem lactose e 30 de leite UHT comum. Os espectros de refletância difusa foram obtidos em um FT-NIR na faixa de 833 a 2500 nm, com resolução de 4 cm^{-1} e média de 16 varreduras, e pelo NIR ultracompacto na faixa de 908 a 1676 nm, com resolução de 12,5 nm e média de 50 varreduras. Os modelos de classificação foram obtidos pelos métodos PLS-DA e LDA com seleção de variáveis robustas por SPA e GA, avaliando-se diferentes pré-tratamento espectrais (MSC, SNV e derivadas). Os três modelos desenvolvidos (PLS-DA, GA-LDA e SPA-LDA) com espectros obtidos no equipamento de bancada classificaram corretamente todas as amostras, com sensibilidade e especificidade de 100%. Para os espectros obtidos no equipamento portátil, com os modelos PLS-DA e GA-LDA obteve-se sensibilidade e especificidade de 100%. Já o modelo SPA-LDA apresentou sensibilidade e especificidade de 80% e 100%, respectivamente. Esses resultados indicam que metodologias utilizando equipamentos NIR, em especial o portátil ultracompacto, com técnicas multivariadas de classificação são viáveis na discriminação entre leites com e sem lactose em campo sendo, portanto, uma alternativa não destrutiva, rápida e precisa.

Palavras-chave: Lactose, NIRS, NIR portátil, classificação multivariada.

Multivariate classification of UHT milk for the presence of lactose using portable and benchtop NIR spectrometers

Abstract: Lactose hydrolyzed milk was developed from the 1970s in order to serve individuals with lactose intolerance. Thus, the demand for lactose-free products by intolerant consumers has created a market segment for this foodstuff whose quality control has to be guaranteed. For the purpose of assess milk samples as regards the lactose content, this work proposes an analytical methodology to classify regular and lactose-free UHT milks using multivariate classification methods and NIR spectra obtained in FT-NIR and ultra-compact NIR spectrometers, aiming at field analysis. For this, a total of 71 samples were purchased, 41 were lactose-free UHT milk and 30 regular UHT milk. Diffuse reflectance spectra were obtained by FT-NIR in the wavelength range of 833 to 2500 nm, 4 cm^{-1} of resolution and a mean of 16 scans, and by the ultracompact NIR in the wavelength range of 908 to 1676 nm, with a resolution of 12,5 nm of resolution and a mean of 50 scans. The classification models were obtained by the PLS-DA and LDA techniques with robust variables selection by SPA and GA, evaluating different spectral pre-processing (MSC, SNV and derivatives). The three models developed (PLS-DA, GA-LDA and SPA-LDA) with spectra obtained in the benchtop equipment correctly classified all the samples with sensitivity and specificity of 100%. For the spectra obtained in the portable equipment, the PLS-DA and GA-LDA models obtained sensitivity and specificity of 100%. On the other hand, the SPA-LDA model presented sensitivity and specificity of 80% and 100%, respectively. These results indicate that methodologies using NIR equipment, especially the ultracompact portable, with multivariate classification techniques are feasible in the discrimination between regular and lactose-free milk in the field, thus becoming a non-destructive, quick and precise alternative analysis.

Key words: Lactose, NIRS, portable NIR, Multivariate classification.

1. Introdução

A lactose (O- β -D-galactopiranosil-(1 \rightarrow 4)- β -D-glucopiranosose, C₁₂H₂₂O₁₁) é um dissacarídeo predominante no leite composto por uma molécula de glicose ligada a uma molécula de galactose. Apresenta-se em equilíbrio nas formas anoméricas α - e β - em soluções aquosas, possui menor solubilidade quando comparada a outros dissacarídeos e doçura próxima a 20% da doçura relativa à sacarose (GÄNZLE *et al.*, 2008; ILLANES, 2016).

Devido a uma deficiência na produção da enzima intestinal lactase em relação à quantidade de lactose ingerida, alguns indivíduos desenvolvem intolerância à lactose, condição caracterizada pela incapacidade de digestão do açúcar do leite (BAHNA, 1996). Para atender a necessidade desses indivíduos, desde os anos 1970, leite sem lactose é produzido geralmente pela aplicação da enzima β -galactosidase que promove a hidrólise da lactose, a partir de processo simples e que não requer equipamentos específicos na linha de produção (MORLOCK *et al.*, 2014; HARJU *et al.*, 2012).

O desenvolvimento de técnicas analíticas mais sofisticadas e acessíveis se faz importante para acompanhar os avanços da indústria moderna. A espectroscopia NIR é uma técnica capaz de fornecer informações valiosas sobre várias propriedades de uma amostra. Para extrair a informação contida no espectro, caracterizado por bandas sobrepostas, é necessário utilizar análise multivariada de dados para minimizar os efeitos que contribuem para a complexidade do espectro como a natureza e heterogeneidade da amostra, ruído instrumental, dispersão e efeitos ambientais, para assim melhorar as características espectrais, remover as fontes de variação indesejadas e alcançar o reconhecimento de padrões desejado (SANTOS *et al.*, 2017).

Vários estudos fundamentam a aplicabilidade da espectroscopia NIR combinada à classificação multivariada para análises de leite, apesar da literatura científica ainda apresentar-se escassa em relação à classificação de leite UHT quanto a presença ou ausência de lactose, principalmente utilizando o NIR ultracompacto.

Lu *et al.* (2009) compararam o desempenho da Análise Discriminante pelo Método dos Quadrados Mínimos Parciais (PLS-DA) com o método máquinas de vetores de suporte por mínimos quadrados (*Least Squares-Support Vector Machine* – LS-SVM) em combinação à espectroscopia NIR para detecção de melamina em leite em pó. Nesse caso,

os resultados do modelo LS-SVM desenvolvido mostraram que todas as amostras do conjunto de treinamento e de previsão foram classificadas corretamente, enquanto o modelo PLS-DA classificou incorretamente 43 amostras no conjunto de treinamento e 38 no conjunto previsão. Uma possível razão encontrada para essa classificação errada foi a existência de relação não linear entre as amostras das duas classes, por isso o LS-SVM obteve melhor desempenho (LU *et al.*, 2009).

Valenti *et al.* (2013) elaboraram um modelo de classificação PLS-DA para discriminação de leites com base em diferentes sistemas de alimentação fornecidos às vacas. Os espectros NIR brutos na faixa de 1100 a 2498 nm foram pré-processados pelo procedimento de padronização normal de sinal com correção da linha de base (*Standard Normal Variate and De-trend – SNVD*) e primeira derivada. Os espectros derivados passaram então pela análise PLS-DA para discriminação entre os tratamentos de alimentação das vacas. As diferenças na composição do leite decorrentes das práticas de alimentação provavelmente foram detectadas nos espectros, permitindo a discriminação entre os sistemas de alimentação com os resultados alinhados aos relatos da literatura, descrevendo a possibilidade de usar espectros NIR com certa eficiência para rastrear sistemas de alimentação (VALENTI *et al.*, 2013).

Mabood *et al.* (2017) investigaram o desenvolvimento de um novo método de espectroscopia NIR combinado a Quimiometria para detectar a adulteração de adição de leites comerciais mais baratos em leite de camelo. O método PLS-DA foi aplicado aos espectros na faixa de 1330 a 2500 nm pré-tratados por primeira derivada de Savitzky-Golay (polinômio de 2ª ordem e janela de 5 pontos) para detectar a discriminação entre as amostras de leite de camelo puro e adulterado. O modelo construído mostrou uma clara distinção entre o leite de camelo puro e o adulterado, sugerindo a espectroscopia NIR mais o método PLS-DA como ferramentas de identificação de adulteração em leite de camelo (MABOOD *et al.*, 2017).

Aliakbarian *et al.* (2015) investigaram a capacidade da espectroscopia NIR e da classificação multivariada empregando Análise Discriminante Linear (LDA) combinada a seleção de variáveis pelo algoritmo *stepwise* de descorrelação na discriminação entre amostras de leite fermentado adicionado de extratos fenólicos obtidos de bagaços de azeitona e uva. Os resultados mostraram uma adequada capacidade de classificação da LDA, provando que a espectroscopia NIR pode ser considerada como alternativa rápida e

confiável para a discriminação entre as amostras de leite fermentado probiótico de acordo com o tipo de extrato adicionado (ALIAKBARIAN *et al.*, 2015).

A espectroscopia no infravermelho próximo (NIR) apresenta vantagens como curto tempo de análise, mínima ou nenhuma preparação da amostra, ser não destrutiva, robusta e de custo relativamente baixo, além da possibilidade de análises em campo utilizando equipamentos portáteis mais práticos e menos onerosos (AYVAZ & RODRIGUEZ-SAONA, 2015; AGELET & HURBURGH JR., 2014).

Os produtos lácteos sem lactose têm conquistado cada vez mais espaço e dessa forma, métodos de análise que permitam um controle efetivo quanto a presença ou ausência de lactose de forma rápida, precisa e de relativo baixo custo tornam-se relevantes (JELEN & TOSSAVAINEN, 2003; BUNACIU *et al.*, 2016). Diante disso, neste estudo foi proposto um método analítico para discriminar leites UHT comum e sem lactose utilizando classificação multivariada e espectros NIR obtidos em espectrômetros FT-NIR de bancada e NIR ultracompacto tendo em vista o monitoramento da presença de lactose nos leites sem lactose.

2. Material e métodos

2.1. Amostras

Um total de 71 amostras de leite UHT foram adquiridas no comércio local do estado de Pernambuco-Brasil. A escolha das amostras considerou diferentes marcas e lotes entre as variedades integral, semidesnatado e desnatado. Das 71 amostras, 41 foram de leite UHT sem lactose e 30 de leite UHT com lactose. Todas as amostras foram armazenadas a temperatura ambiente, mantidas nas embalagens originais fechadas e analisadas sem qualquer pré-tratamento prévio.

2.2. Aquisição dos espectros NIR

As medidas espectrométricas no modo de refletância difusa foram realizadas nos espectrômetros Spectrum Frontier FT-NIR da marca Perkin Elmer (equipamento de bancada de alta resolução) e MicroNIRTM Spectrometer 1700 da marca JDSU (equipamento ultracompacto de baixa resolução). As leituras obtidas do espectrômetro Spectrum Frontier FT-NIR foram realizadas utilizando-se o acessório NIRA *reflectance* na faixa de 833 a 2500 nm, com resolução de 4 cm⁻¹, média de 16 varreduras e incremento espectral de 0,2 nm, totalizando 8003 variáveis. As leituras adquiridas no espectrômetro NIR ultracompacto foram efetuadas na faixa de 908 a 1676 nm com uma resolução de 12,5 nm, média de 50 varreduras, e incremento espectral de 6,2 nm, totalizando 125 variáveis. O padrão de politetrafluoretileno espectroscópico foi utilizado como branco em ambos os equipamentos.

As amostras de leites UHT com e sem lactose foram dispostas em célula de vidro em quantidade suficiente (aproximadamente 2 mL) para formação do filme pela sobreposição da peça de metal, sem pré-preparo e imediatamente após a abertura das embalagens. As leituras de cada amostra foram realizadas em sequência nos equipamentos de bancada e ultracompacto. Água destilada e acetona P.A. foram utilizadas para limpeza das células entre as leituras. A aquisição dos espectros foi realizada em sala climatizada, com temperatura média de 23 °C e umidade relativa de 52%.

2.3. Tratamento dos dados

Os espectros NIR centrados na média foram pré-processados a fim de eliminar informações irrelevantes tais como espalhamento da radiação. Diferentes técnicas de pré-processamento foram avaliadas, dentre elas, *Multiplicative Scatter Correction* (MSC), *Standard Normal Variate* (SNV) e primeira e segunda derivadas por filtros Savitzky-Golay (SG) com diferentes tamanhos de janela.

Os métodos de classificação por Análise Discriminante pelo Método dos Quadrados Mínimos Parciais (*Partial Least Squares-Discriminant Analysis* – PLS-DA) e Análise Discriminante Linear (*Linear Discriminant Analysis* – LDA) com seleção de variáveis pelo

Algoritmo Genético (*Genetic Algorithm* – GA) e Algoritmo das Projeções Sucessivas (*Successives Projections Algorithm* – SPA) foram utilizados na construção dos modelos de classificação.

Para cada conjunto de espectros, classe de leite (com e sem lactose) e método de classificação foram selecionadas amostras de treinamento separadamente, utilizando o algoritmo Kennard-Stone, para contornar a seleção desigual de amostras entre as classes de leite e garantir a representatividade dos conjuntos (PONTES *et al.*, 2005). Uma análise de componentes principais (PCA) foi realizada preliminarmente para verificar o comportamento das amostras e presença de *outliers*.

O modelo PLS-DA desenvolvido utilizando 70% das amostras para o conjunto de treinamento, 30% para o conjunto de previsão e o procedimento de validação cruzada *leave-one-out* para seleção do melhor número de variáveis latentes. As amostras para os modelos SPA-LDA e GA-LDA foram distribuídas entre os conjuntos de treinamento, validação e previsão em 70%, 14% e 16%, respectivamente.

No modelo PLS-DA, onde uma regressão clássica por mínimos quadrados parciais procura variáveis latentes que apresentem máxima covariância com a variável dependente (classes), as amostras foram divididas em duas classes, com lactose e sem lactose, e categorizadas utilizando 0 e 1 (BORRÀS *et al.*, 2014; MARQUETTI *et al.*, 2016). A validação cruzada *leave-one-out* foi empregada para selecionar o número de variáveis latentes. O valor limite (*threshold*) aplicado para os modelos PLS-DA foi 0,5. A detecção de *outliers* foi realizada observando os gráficos de *scores*, resíduos e influência (*leverage*) da PCA.

Nos modelos obtidos por LDA, onde se utiliza a distância de Mahalanobis para encontrar funções discriminantes lineares que maximizam a variância entre as classes e minimizam a variância dentro da classe, sendo o número de variáveis espectrais grande, o emprego de técnicas de seleção de variáveis mostra-se essencial (MOREIRA *et al.*, 2009; HUANG *et al.*, 2016). Neste caso, algoritmo genético (GA) e o algoritmo das projeções sucessivas (SPA) foram avaliados (SILVA *et al.*, 2013).

Na seleção de variáveis utilizando SPA as amostras de treinamento e validação foram empregadas para selecionar o melhor subconjunto de comprimentos de onda com base na função de custo médio de erro de classificação por LDA. O conjunto de previsão

com amostras independentes foi usado para verificar a capacidade de generalização do classificador resultante na avaliação final (PONTES *et al.*, 2011).

Para a seleção de variáveis utilizando GA foram realizadas 100 gerações partindo de 200 indivíduos. As taxas de mutação e cruzamento foram estabelecidas pelo algoritmo em 10% e 60%, respectivamente. A rotina foi repetida 5 vezes e o modelo com menor valor da função de custo e menores taxas de erros foi utilizado para a seleção de variáveis pelo algoritmo genético (BRITO *et al.*, 2013).

Os modelos PLS-DA e LDA foram avaliados com base na complexidade do modelo e o critério de seleção foi fundamentado no que apresentou melhor sensibilidade e especificidade, obtidas de acordo com as Equações 1 e 2 (OLIVERI & DOWNEY, 2012), combinado ao menor número de variáveis latentes para o PLS-DA que conduziu ao modelo mais parcimonioso.

$$\text{sensibilidade} = \frac{VP}{VP + FN} \quad (1)$$

$$\text{especificidade} = \frac{VN}{VN + FP} \quad (2)$$

Onde, para a sensibilidade da classificação, o parâmetro verdadeiro positivo (VP) correspondeu às amostras com lactose classificadas corretamente e o parâmetro falso negativo (FN) foi definido pelas amostras com lactose classificadas erroneamente. Para a especificidade da classificação, o parâmetro verdadeiro negativo (VN) foi determinado pelas amostras sem lactose corretamente classificadas e o parâmetro falso positivo (FP) correspondeu às amostras sem lactose classificadas incorretamente.

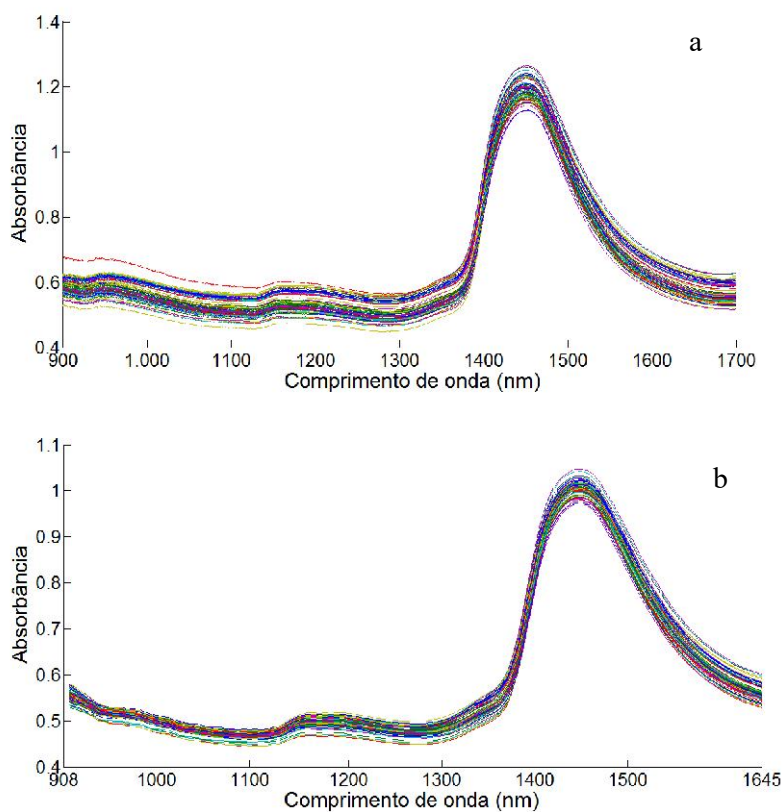
O pacote computacional The Unscrambler® v9.7 (CAMO S.A.) foi utilizado para organização da matriz de dados e pré-processamento de amostras e de variáveis. O programa computacional Matlab® R2010a 7.10.0.499 (Mathworks) e PLSToolbox 8.21 Demo (Eigenvector Research, Inc.) foram utilizados para seleção dos conjuntos de treinamento e teste pelo algoritmo Kennard-Stone (KENNARD & STONE, 1969), seleção de variáveis GA e SPA, e para construção dos modelos de classificação PLS-DA e LDA.

3. Resultados e discussão

3.1. Espectros NIR nos equipamento de bancada e portátil

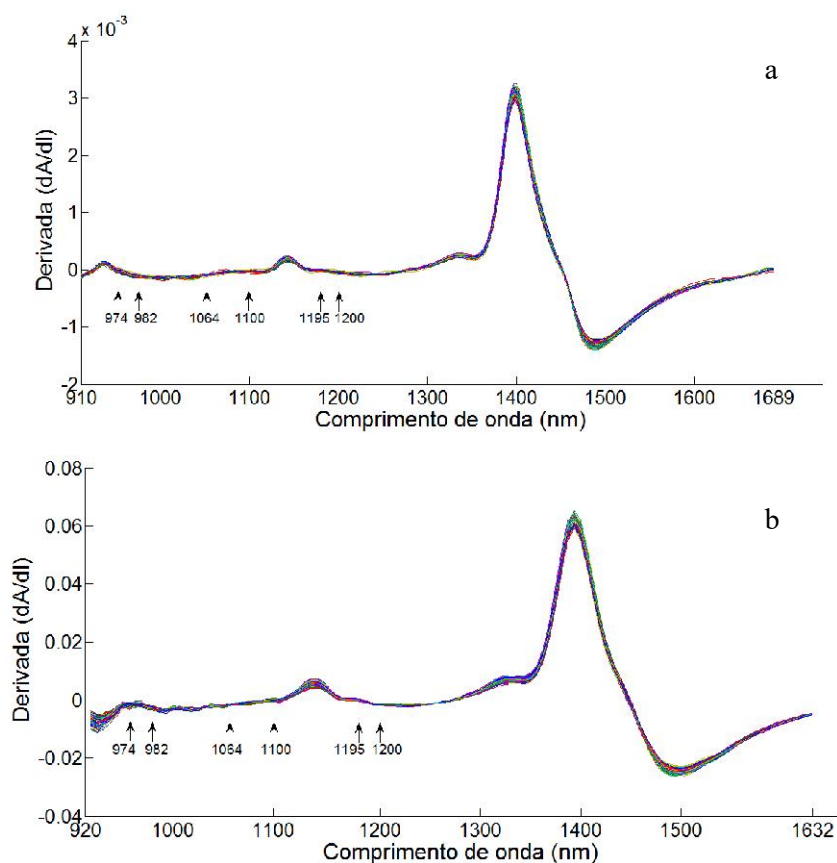
Os espectros obtidos no equipamento de bancada (FT-NIR) foram medidos até 2500 nm, entretanto, para efeitos de comparação com os espectros obtidos no equipamento ultracompacto foi dado ênfase a região entre 900 a 1700 nm. A Figura 1 apresenta os espectros brutos obtidos no FT-NIR de bancada (Figura 1a) na faixa de 833 a 1700 nm, e no NIR ultracompacto (Figura 1b) na faixa de 908 a 1676 nm para as amostras de leite com e sem lactose. Os espectros apresentam uma banda dominante ao redor de 1450 nm, devido à combinação de estiramentos simétricos e assimétricos, relacionada ao primeiro sobreton do grupo O-H da água, além do segundo e terceiro sobretons de ligações C-H entre 1100-1200 nm e entre 900-1000 nm, respectivamente.

Figura 1. Espectros NIR com e sem lactose brutos obtidos com equipamentos FT-NIR de bancada (a) e NIR ultracompacto (b).



Os espectros apresentam ainda efeitos aditivos na linha de base, devidos ao espalhamento da radiação, que foram melhor corrigidos pela primeira derivada com filtros Savitzky-Golay com polinômio de 2ª ordem e janelas de 105 e 5 pontos para os espectros resultantes do FT-NIR e do NIR ultracompacto, respectivamente (Figura 2). O número elevado de pontos da janela na etapa de suavização pelos filtros Savitzky-Golay foi devido ao incremento espectral (0,2 nm) utilizado na aquisição dos espectros no FT-NIR, o que gerou um espectro com 8003 variáveis.

Figura 2. Espectros NIR com e sem lactose pré-processados por 1ª derivada com filtros Savitzky-Golay obtidos nos equipamentos de bancada (a) e ultracompacto (b).



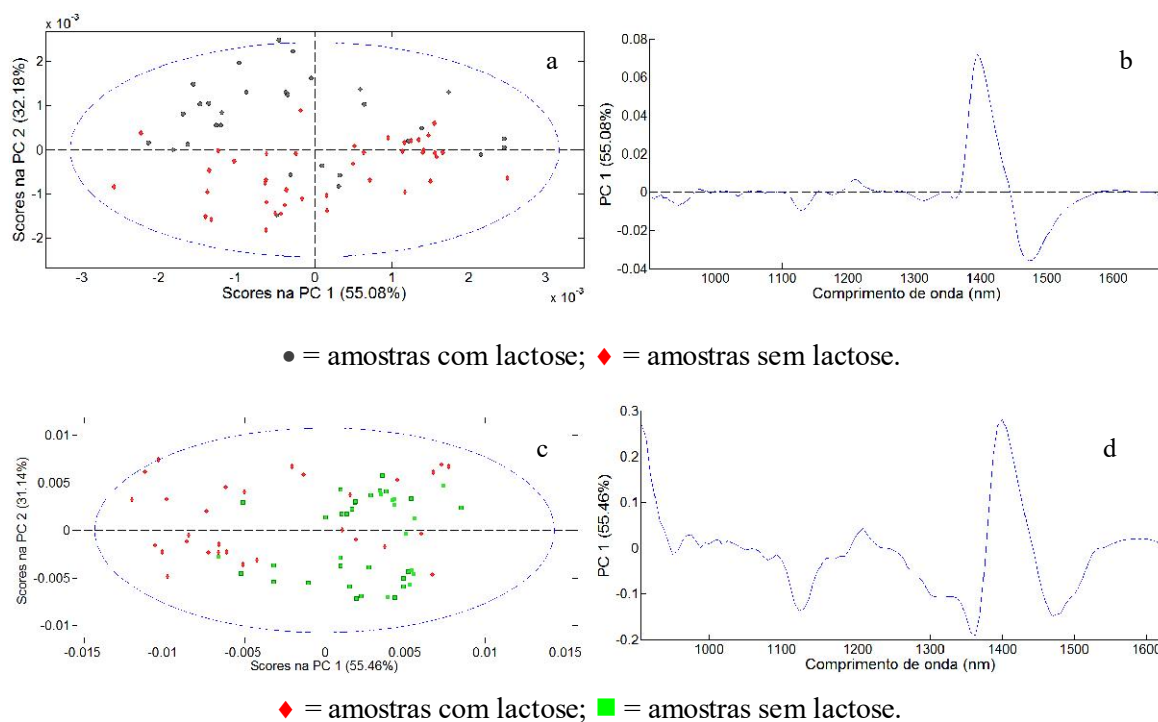
As informações relacionadas às bandas de absorções dos grupos funcionais O-H e C-H relevantes para o conteúdo de lactose foram evidenciadas pela primeira derivada aplicada aos espectros brutos obtidos em ambos os equipamentos, bem como informação relevante sobre a lactose em 974, 982 e 1064 nm em destaque na Figura 2 (TSENKOVA *et*

al., 1999). Informações referentes ao segundo sobretom de ligações C-H diferentes também podem ser encontradas em torno de 1195 nm (MELFSEN *et al.*, 2012). Outras bandas atribuídas às vibrações de OH e CH para açúcares relatadas na literatura foram relacionadas ao segundo sobretom de vibrações OH em torno de 1000 nm, sobretoms da combinação OH em torno de 1100 nm e ao segundo sobretom das bandas de estiramento do grupo CH entre 1100 e 1200 nm (GOLIC *et al.*, 2003).

A PCA nos espectros FT-NIR e NIR ultracompacto na região entre 900 a 1700 nm pré-processados não apresentou nenhuma separação efetiva entre as classes com e sem lactose. As Figuras 3a e 3c apresentam os gráficos dos escores da PC1xPC2 para o FT-NIR e NIR ultracompacto, respectivamente, onde se observa as amostras dispersas e sobrepostas entre as classes com e sem lactose. Para o FT-NIR PC1 e PC2 explicam 55,08% e 32,18% da variação dos dados, e para o NIR ultracompacto 55,46% e 31,14% respectivamente.

Amostras posicionadas fora da elipse a 95% de confiança, potenciais outliers, foram avaliadas considerando os valores de *leverage* e resíduos de Student e essa influência foi reavaliada durante o desenvolvimento dos modelos de classificação. Nos gráficos dos pesos (Figuras 3b e 3d) para PC1 predomina a influência das variáveis em torno de 1450 nm, referente ao primeiro sobretom do grupo OH da água, e da região entre 1100 e 1200 nm, correspondente ao segundo sobretom das bandas de estiramento do grupo CH atribuídas aos açúcares.

Figura 3. PCA dos espectros com e sem lactose obtidos com equipamentos FT-NIR (a e b) e NIR ultracompacto (c e d) pré-processados por 1ª derivada com filtro SG. (a) e (c) gráfico dos escores; (b) e (d) gráficos dos pesos.



3.2. Modelos de classificação

Os modelos de classificação PLS-DA e LDA foram desenvolvidos a partir dos espectros pré-processados pela primeira derivada com filtros Savitzky-Golay com polinômio de 2ª ordem e janela de 105 e 5 pontos para os equipamentos FT-NIR e NIR ultracompacto, respectivamente, na região entre 900 a 1700 nm.

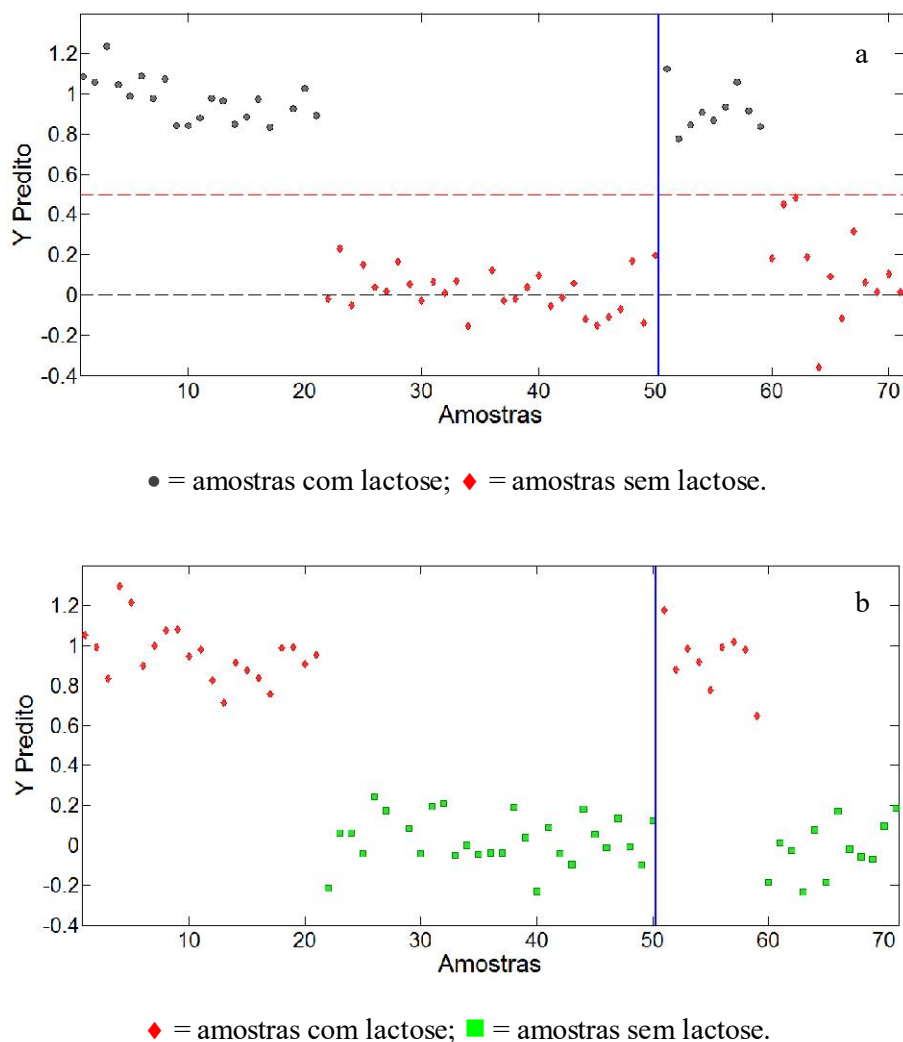
Para os espectros obtidos no FT-NIR, o modelo PLS-DA foi obtido com 48 amostras, pois 2 *outliers* foram identificados e removidos para aprimorar o conjunto de treinamento. As amostras com e sem lactose foram categorizadas por 0 e 1, respectivamente, e o valor limite (*threshold*) de 0,5, onde amostras com valores limites abaixo deste limite são classificadas como sem lactose e com valores limites acima de 0,5 amostras com lactose.

A validação cruzada mostrou, com sete variáveis latentes, sensibilidade de 100% para a classe com lactose e 96,4% para a classe sem lactose, e especificidade de 96,4% e 100% para as classes com e sem lactose, respectivamente. Ainda assim, foi observado que

todas as amostras foram classificadas corretamente pelo modelo desenvolvido para os espectros obtidos no FT-NIR, que mostrou sensibilidade e especificidade de 100% para os conjuntos de treinamento e previsão. A Figura 4a apresenta os resultados nas etapas de treinamento e previsão pelo modelo obtido.

Para os espectros obtidos no NIR ultracompacto o modelo PLS-DA foi obtido utilizando 49 amostras no conjunto de treinamento, onde se identificou apenas uma amostra *outlier*. Todas as amostras foram classificadas corretamente e o modelo desenvolvido utilizando cinco variáveis latentes mostrou sensibilidade e especificidade de 100% (Figura 4b). Comparado ao modelo PLS-DA para os espectros do FT-NIR de bancada, embora ambos apresentem classificação correta de todas as amostras de previsão, o modelo obtido com NIR ultracompacto parece discriminar melhor as amostras com e sem lactose, visto que essas estão mais distantes do valor limite (0,5).

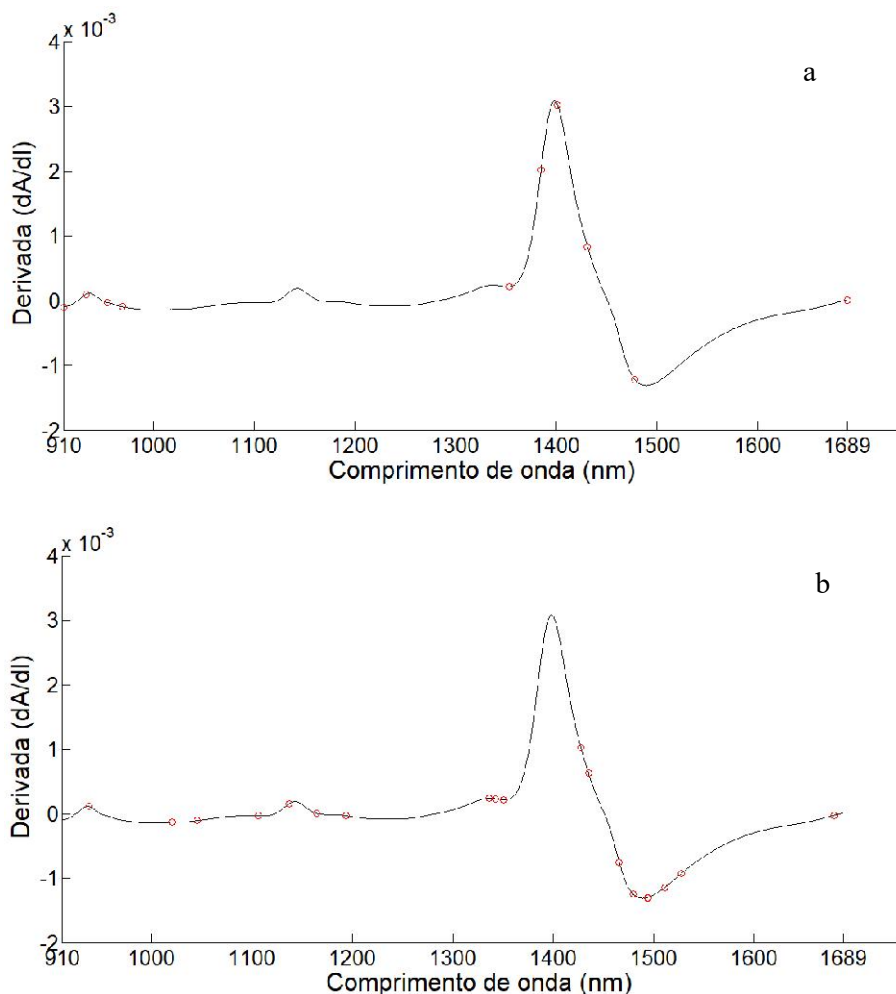
Figura 4. Resultados das etapas de treinamento e previsão para o modelo PLS-DA a partir dos espectros do FT-NIR (a) e NIR ultracompacto (b). A esquerda da linha azul estão as amostras de treinamento e a direita, amostras de previsão.



O modelo SPA-LDA para os espectros FT-NIR foi obtido a partir de 10 variáveis selecionadas conforme Figura 5a. Esse modelo classificou todas as amostras de previsão do leite UHT corretamente com sensibilidade e a especificidade de 100%. As variáveis selecionadas pelo SPA têm informação nas regiões de absorção da lactose e de grupos OH.

Para obtenção do modelo GA-LDA a partir dos espectros FT-NIR foram selecionadas 19 variáveis informativas pelo algoritmo GA (Figura 5b). As variáveis selecionadas são referentes às absorções de lactose e ligações CH. Esse modelo discriminou todas as amostras de previsão em suas respectivas classes, com sensibilidade e a especificidade de 100%.

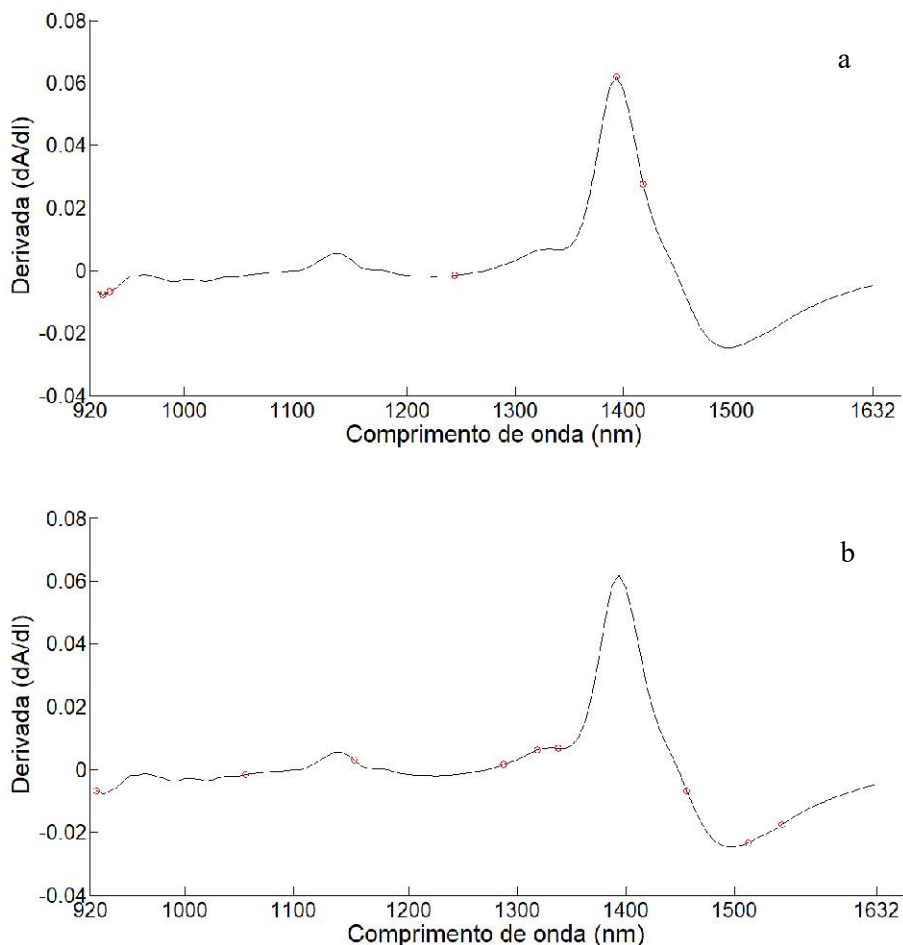
Figura 5. Espectro médio derivativo obtido no FT-NIR com variáveis selecionadas pelo SPA (a) e GA (b).



Considerando os modelos LDA a partir dos espectros do NIR ultracompacto, o modelo SPA-LDA foi obtido com 5 variáveis selecionadas pelo SPA (Figura 6a) com informação de absorção da lactose, mas também do grupo OH. Esse modelo apresentou sensibilidade e a especificidade de 80% e 100%, respectivamente, que se deve a uma única amostra da classe com lactose erroneamente classificada na classe das amostras sem lactose.

O modelo GA-LDA para os espectros do NIR ultracompacto selecionou 9 variáveis (Figura 6b), que assim como para os espectros do FT-NIR, as variáveis são representativas das absorções da lactose e de grupos CH. Esse modelo também classificou todas as amostras corretamente, apresentando sensibilidade e a especificidade de 100%.

Figura 6. Espectro médio derivativo obtido no NIR ultracompacto com variáveis selecionadas pelo SPA (a) e GA (b).



Possivelmente, a distinção entre as duas classes foi influenciada pela presença de glicose resultante do processo de hidrólise no leite sem lactose, que pode ser observada pela seleção dos comprimentos de onda próximos a 1195, 1385, 1520, 1688 nm, relacionados às bandas de absorção de glicose (GIANGIACOMO *et al.*, 1981; GOSH & ROY, 1988). A presença dos monossacarídeos glicose e galactose, do dissacarídeo lactose e outros oligossacarídeos produtos do processo de hidrólise presentes no leite também podem ser consideradas na discriminação das classes, como demonstra a seleção de variáveis próximas a 1432 e 1457 nm, comprimentos de ondas mais característicos para oligossacarídeos e monossacarídeos, respectivamente (ROBERT & CADET, 1998). A tabela de contingência (Tabela 1) resume os resultados dos modelos PLS-DA, SPA-LDA e GA-LDA obtidos para ambos os equipamentos.

Tabela 1. Tabela de contingência para os modelos de classificação dos espectros NIR de bancada e ultracompacto.

Equipamento de bancada (FT-NIR)									
Conjunto de previsão		PLS-DA (7)		SPA-LDA		GA-LDA			
		Classes previstas		Classes previstas		Classes previstas			
Classes verdadeiras	n	0	1	n	0	1	n	0	1
0	9	9	-	5	5	-	5	5	-
1	12	-	12	6	-	6	6	-	6
Classificação correta		100%			100%			100%	

Equipamento portátil (NIR ultracompacto)									
Conjunto de previsão		PLS-DA (5)		SPA-LDA		GA-LDA			
		Classes previstas		Classes previstas		Classes previstas			
Classes verdadeiras	n	0	1	n	0	1	n	0	1
0	9	9	-	5	4	1	5	5	-
1	12	-	12	6	-	6	6	-	6
Classificação correta		100%			80%			100%	

0 = com lactose; 1 = sem lactose; n = número de amostras do conjunto de previsão.

Alguns estudos empregaram a espectroscopia NIR e classificação multivariada para detecção de adulterantes em leite em pó (LU *et al.*, 2009) e leite de camelo (MABOOD *et al.*, 2017), discriminação de leites com base em diferentes sistemas de alimentação (VALENTI *et al.*, 2013) e discriminação entre amostras de leite fermentado adicionado de extratos fenólicos (ALIAKBARIAN *et al.*, 2015), por exemplo. Entretanto, espectroscopia NIR associada à classificação multivariada para avaliação de leite UHT quanto à presença de lactose na literatura científica apresenta-se escassa.

3.3. Comparação entre os modelos de classificação e os equipamentos

A espectroscopia NIR, na faixa espectral entre 900 e 1700 nm, associada a métodos de classificação multivariada (PLS-DA, SPA-LDA e GA-LDA) demonstrou ser capaz de

distinguir amostras de leite com e sem lactose. Para ambos os equipamentos (FT-NIR e NIR ultracompacto) os três modelos apresentaram bom desempenho discriminante para a lactose com os modelos SPA-LDA e GA-LDA apresentando-se como modelos mais simples e parcimoniosos que o PLS-DA devido ao uso de um menor número das variáveis espectrais originais (FERREIRA, 2015).

Embora o modelo SPA-LDA para o NIR ultracompacto tenha apresentado um desempenho menos eficiente comparado aos modelos PLS-DA e GA-LDA, os resultados foram comparáveis aos resultados obtidos pelos modelos do FT-NIR, equipamento que apresenta maior resolução e precisão. A Tabela 2 apresenta para fins de comparação, os resultados de desempenho dos modelos e dos equipamentos utilizados.

Tabela 2. Desempenho dos modelos de classificação obtidos pelo FT-NIR e NIR ultracompacto.

Equipamento de bancada (FT-NIR)			
	Sensibilidade	Especificidade	Variáveis latentes
PLS-DA	100%	100%	7
			Variáveis selecionadas
SPA-LDA	100%	100%	10
GA-LDA	100%	100%	19
Equipamento NIR ultracompacto			
	Sensibilidade	Especificidade	Variáveis latentes
PLS-DA	100%	100%	5
			Variáveis selecionadas
SPA-LDA	80%	100%	5
GA-LDA	100%	100%	9

Os métodos empregados no presente trabalho foram baseados em medidas espectroscópicas e classificação multivariada, indicando uma boa resolutividade para a distinção entre os tipos de leite com e sem lactose em equipamento de bancada e portátil.

Além disso, uma seleção de variáveis apropriada fornece informações úteis que podem colaborar com informações para o desenvolvimento de modelos empregando equipamentos portáteis.

Apesar do bom desempenho de ambos os equipamentos na distinção entre amostras com e sem lactose, a pequena dimensão do NIR ultracompacto (45 mm de diâmetro x 42 mm de altura) e o reduzido peso (cerca de 60g) permitem que esse seja facilmente manuseado possibilitando seu uso em campo, o que é muito vantajoso em ações de fiscalização.

4. Conclusão

Este trabalho apresentou uma metodologia rápida e precisa para discriminação entre leites UHT com e sem lactose a partir de medidas espectroscópicas na região NIR aliadas a métodos de classificação multivariada (PLS-DA e LDA) com dois equipamentos diferentes, o FT-NIR de bancada de alta resolução e o NIR ultracompacto de baixa resolução. Os equipamentos de bancada e portátil apresentaram ótimas sensibilidade e especificidade para os modelos de classificação, com exceção do modelo SPA-LDA desenvolvido para o NIR ultracompacto. Além disso, tanto PLS-DA, método que utiliza de todas as variáveis, quanto LDA associado a técnicas de seleção de variáveis informativas, apresentaram excelente desempenho na classificação das amostras, sendo este último uma alternativa mais parcimoniosa. Dessa forma, o método desenvolvido se constitui em uma ótima ferramenta para discriminação entre os leites UHT com e sem lactose, principalmente, ao empregar equipamento portátil que favorece a avaliação em campo e contribui para aumentar o nível tecnológico com testes sensíveis e específicos mais rápidos e não destrutivos.

5. Referências

- 1) AGELET, L. E.; HURBURGH JR., C. R. Limitations and current applications of Near Infrared Spectroscopy for single seed analysis. **Talanta**. v.121, p.288–299, abr. 2014.

- 2) ALIAKBARIAN, B.; CASALE, M.; PAINI, M.; CASAZZA, A. A.; LANTERI, S.; PEREGO, P. Production of a novel fermented Milk fortified with natural antioxidants and its analysis by NIR spectroscopy. **LWT – Food Science and Technology**. v.62, n.1, p.376–383, jun. 2015.
- 3) AYVAZ, H.; RODRIGUEZ-SAONA, L. E. Application of handheld and portable spectrometers for screening acrylamide content in commercial potato chips. **Food Chemistry**. v.174, p.154–162, mai. 2015.
- 4) BAHNA, S. L. Is it milk allergy or lactose intolerance? **Immunology and Allergy Clinics of North America**. v.16, n.1, p.187-198, fev. 1996.
- 5) BORRÀS, E.; AMIGO, J. M.; BERG, F. V. D.; BOQUÉ, R.; BUSTO, O. Fast and robust discrimination of almonds (*Prunus amygdalus*) with respect to their bitterness by using near infrared and partial least squares-discriminant analysis. **Food Chemistry**. v.153, p.15–19, jun. 2014.
- 6) BRITO, A. L. B.; BRITO, L. R.; HONORATO, F. A.; PONTES, M. J. C.; PONTES, L. F. B. L. Classification of cereal bars using near infrared spectroscopy and linear discriminant analysis. **Food Research International**. v.51, n.2, p.924–928, mai. 2013.
- 7) BUNACIU, A. A.; ABOUL-ENEIN, H. Y.; HOANG, V. D. Vibrational spectroscopy used in milk products analysis: A review. **Food Chemistry** v.196, p.877–884, abr. 2016.
- 8) FERREIRA, M. M. C. Calibração – Métodos de regressão. In: Ferreira, M. M. C. Quimiometria – conceitos, métodos e aplicações / Márcia Miguel Castro Ferreira – Campinas, SP: **Editora da Unicamp**. 496p., p.251–404, 2015.
- 9) GÄNZLE, M. G.; HAASE, G.; JELEN, P. Lactose: Crystallization, hydrolysis and value-added derivatives. **International Dairy Journal**. v.18, n.7, p.685–694, jun. 2008.
- 10) GIANGIACOMO, R.; MAGEE, J. B.; BIRTH, G. S.; DULL, G. G. Predicting concentrations of individual sugars in dry mixtures by near-infrared reflectance spectroscopy. **Journal of Food Science**. v.46, n.2, p.531–534, mar. 1981.
- 11) GOLIC, M.; WALSH, K.; LAWSON, P. Short-Wavelength Near-Infrared Spectra of Sucrose, Glucose, and Fructose with Respect to Sugar Concentration and Temperature. **Applied Spectroscopy**. v.57, n.2, p.139–145, fev. 2003.
- 12) HARJU, M.; KALLIOINEN, H.; TOSSAVAINEN, O. Lactose hydrolysis and other conversions in dairy products: Technological aspects. **International Dairy Journal** v.22, n.2, p.104-109, fev. 2012.

- 13) HUANG, Y.; WU, Z.; SU, R.; RUAN, G.; DU, F.; LI, G. Current application of chemometrics in traditional Chinese herbal medicine research. **Journal of Chromatography B**. v.1026, p.27–35, jul. 2016.
- 14) ILLANES, A. Lactose: Production and Upgrading. In: Illanes, A.; Guerrero, C.; Vera, C.; Wilson, L.; Conejeros, R.; Scott, F. Lactose-Derived Prebiotics - A Process Perspective. **Academic Press**. 312p., p.1-33, 2016.
- 15) KENNARD, R. W.; STONE, L. A. Computer aided design of experiments. **Technometrics**. v.11, n.1, p.137–148, 1969.
- 16) LU, C.; XIANG, B.; HAO, G.; XU, J.; WANG, Z.; CHEN, C. Rapid detection of melamine in milk powder by near infrared spectroscopy. **J. Near Infrared Spectrosc.** v.17, p.59–67, mar. 2009.
- 17) MABOOD, F.; JABEEN, F.; AHMED, M.; HUSSAIN, J.; MASHAYKHI, S. A. A. AL; RUBAIEY, Z. M. A. AL; FAROOQ, S.; BOQUÉ, R.; ALI, L.; HUSSAIN, Z.; AL-HARRASI, A.; KHAN, A. L.; NAUREEN, Z.; IDREES, M.; MANZOOR, S. Development of new NIR-spectroscopy method combined with multivariate analysis for detection of adulteration in camel milk with goat milk. **Food Chemistry**. v.221, p.746–750, abr. 2017.
- 18) MARQUETTI, I.; LINK, J. V.; LEMES, A. L. G.; SCHOLZ, M. B. S.; VALDERRAMA, P.; BONA, E. Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee. **Computers and Electronics in Agriculture**. v.121, p.313–319, fev. 2016.
- 19) MELFSEN, A.; HARTUNG, E.; HAEUSSERMANN, A. Accuracy of milk composition analysis with near infrared spectroscopy in diffuse reflection mode. **Biosystems Engineering**. v.112, n.3, p.210–217, jul. 2012.
- 20) MORLOCK, G. E.; MORLOCK, L. P.; LEMO, C. Streamlined analysis of lactose-free dairy products. **Journal of Chromatography A**. v.1324, p.215-223, jan. 2014.
- 21) PONTES, M. J. C.; GALVÃO, R. K. H.; ARAÚJO, M. C. U.; MOREIRA, P. N. T.; NETO, O. D. P.; JOSÉ, G.E.; SALDANHA, T. C. B. The successive projections algorithm for spectral variable selection in classification problems. **Chemometrics and Intelligent Laboratory Systems**. v.78, n.1–2, p.11–18, jul. 2005.
- 22) PONTES, M. J. C.; PEREIRA, C. F.; PIMENTEL, M. F.; VASCONCELOS, F. V. C.; SILVA, A. G. B. Screening analysis to detect adulteration in diesel/biodiesel

- blends using near infrared spectrometry and multivariate classification. **Talanta**. v.85, n.4, p.2159–2165, set. 2011.
- 23) ROBERT, C.; CADET, F. Analysis of near-infrared spectra of some carbohydrates. **Applied Spectroscopy Reviews**. v.33, n.3, p.253–266, ago. 1998.
- 24) SANTOS, C. A. T. DOS; PÁSCOA, R. N. M. J.; LOPES, J. A. A review on the application of vibrational spectroscopy in the wine industry: from soil to bottle. **TrAC - Trends in Analytical Chemistry**. v.88, p.100–118, mar. 2017.
- 25) SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**. v.36, n.8, p.1627-1639, jul. 1964.
- 26) SILVA, C. S.; BORBA, F. S. L.; PIMENTEL, M. F.; PONTES, M. J. C.; HONORATO, R. S.; PASQUINI, C. Classification of blue pen ink using infrared spectroscopy and linear discriminant analysis. **Microchemical Journal**. v.109, p.122–127, jul. 2013.
- 27) TSENKOVA, R.; ATANASSOVA, S.; TOYODA, K.; OZAKI, Y.; ITOH, K.; FEARN, T. Near-Infrared Spectroscopy for Dairy Management: Measurement of Unhomogenized Milk Composition. **J. Dairy Sci**. v.82, n.11, p.2344–2351, nov. 1999.
- 28) VALENTI, B.; MARTIN, B.; ANDUEZA, D.; LEROUX, C.; LABONNE, C.; LAHALLE, F.; LARROQUE, H.; BRUNSCHWIG, P.; LECOMTE, C.; BROCHARD, M.; FERLAY, A. Infrared spectroscopic methods for the discrimination of cows' milk according to the feeding system, cow breed and altitude of the dairy farm. **International Dairy Journal**. v.32, n.1, p.26–32, set. 2013.