

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E
TECNOLOGIA DE ALIMENTOS

ELAINY VIRGINIA DOS SANTOS PEREIRA

NOVAS ESTRATÉGIAS ANALÍTICAS BASEADAS EM
ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO E
IMAGENS DIGITAIS PARA IDENTIFICAÇÃO E
QUANTIFICAÇÃO DE ADULTERAÇÕES EM LEITE CAPRINO

JOÃO PESSOA

2020

ELAINY VIRGINIA DOS SANTOS PEREIRA

NOVAS ESTRATÉGIAS ANALÍTICAS BASEADAS EM
ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO E
IMAGENS DIGITAIS PARA IDENTIFICAÇÃO E
QUANTIFICAÇÃO DE ADULTERAÇÕES EM LEITE CAPRINO

JOÃO PESSOA

2020

ELAINY VIRGINIA DOS SANTOS PEREIRA

NOVAS ESTRATÉGIAS ANALÍTICAS BASEADAS EM
ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO E
IMAGENS DIGITAIS PARA IDENTIFICAÇÃO E
QUANTIFICAÇÃO DE ADULTERAÇÕES EM LEITE CAPRINO

Tese apresentada ao Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos, Centro de Tecnologia, Universidade Federal da Paraíba, em cumprimento aos requisitos para obtenção do título de Doutor em Ciência e Tecnologia de Alimentos.

ORIENTADOR: Dra. Maria Inês Sucupira Maciel

CO-ORIENTADOR: Dr. Paulo Henrique Gonçalves Dias Diniz

JOÃO PESSOA

2020

Catálogo na publicação
Seção de Catalogação e Classificação

P436n Pereira, Elainy Virginia dos Santos.

Novas estratégias analíticas baseadas em espectroscopia no infravermelho próximo e imagens digitais para identificação e quantificação de adulterações em leite caprino / Elainy Virginia dos Santos Pereira. - João Pessoa, 2020.

155 f. : il.

Orientação: Maria Inês Sucupira Maciel.

Coorientação: Paulo Henrique Gonçalves Dias Diniz.

Tese (Doutorado) - UFPB/CT.

1. Tecnologia de alimentos. 2. Autenticidade. 3. Produtos lácteos. 4. NIR portátil. 5. Calibração multivariada. 6. Reconhecimento de padrões. 7. Algoritmo das projeções sucessivas. I. Maciel, Maria Inês Sucupira. II. Diniz, Paulo Henrique Gonçalves Dias. III. Título.

UFPB/BC

CDU 664(043)

ELAINY VIRGINIA DOS SANTOS PEREIRA

NOVAS ESTRATÉGIAS ANALÍTICAS BASEADAS EM ESPECTROSCOPIA NO
INFRAVERMELHO PRÓXIMO E IMAGENS DIGITAIS PARA IDENTIFICAÇÃO E
QUANTIFICAÇÃO DE ADULTERAÇÕES EM LEITE CAPRINO

Tese aprovada em _____ de _____ de _____

BANCA EXAMINADORA

Prof^ª. Dr^ª. Maria Inês Sucupira Maciel –PPGCTA/CT/UFPB

Coordenadora da Banca Examinadora

Prof^ª. Dr^ª Rita de Cassia Ramos do Egypto Queiroga –PPGCTA/CT/UFPB

Examinadora Interna

Prof. Dr. Fábio Anderson Pereira da Silva –PPGCTA/CT/UFPB

Examinador Interno

Prof^ª. Dr^ª Eliane Rolim Florentino –PPGCF/UEPB

Examinadora Externa

Prof. Dr José Germano Vêras Neto –PPGQ/UEPB

Examinador Externo

À minha mãe, ao meu marido e ao meu tio Nilson (*in memoriam*),

Dedico.

AGRADECIMENTOS

Ao término desta jornada, não poderia deixar de agradecer:

A Deus, por ser base tão sólida em minha vida, permitindo assim, que eu trilhe caminhos de crescimento e alcance meus objetivos. “Até aqui nos ajudou o Senhor!!!”;

À minha orientadora, Maria Inês Sucupira Maciel, pela confiança e pelo conhecimento compartilhado;

Ao meu orientador e amigo, Paulo Diniz, por tudo que ensinou e compartilhou nesses anos. Em especial pela paciência e por ultrapassar a relação acadêmica, contribuindo valiosamente não só para minha formação profissional, como também para a minha construção como ser humano;

A UEPB, por permitir meu afastamento integral para capacitação e dar suporte a este projeto;

Ao professor Mário Ugulino, por todas as contribuições para realização deste trabalho;

À minha mãe, Ineide Pereira, por ser meu referencial, meu apoio e meu abrigo por toda a vida. “Mãe, você é o melhor anjo da guarda que alguém poderia desejar. Que sorte a minha em tê-la! Mil vidas não são suficientes para agradecer. Amo-te!”;

Ao meu marido, companheiro de todas as horas, Helder Melo, que apoia física e emocionalmente todas as minhas empreitadas, permanecendo firme, sem me deixar fraquejar. Minha gratidão eterna, amor!;

Aos meus irmãos e sobrinhos, que acompanham de perto todos os meus passos, tropeços e conquistas. Sem o apoio de vocês, essa realização seria impossível;

A toda minha família, em especial a minha avó Dina e ao meu tio Nilson (*in memoriam*), que sempre acreditaram que eu alcançaria o tão sonhado título de Doutora, e para isso ofereceram gratuita e amorosamente todo o suporte necessário;

Aos professores do Programa de Pós Graduação em Ciência e Tecnologia de Alimentos da UFPB, que sempre estiveram dispostos a compartilhar seus conhecimentos, contribuindo de forma valiosa para minha formação;

Às minhas amigas e companheiras de Doutorado: Aline, Camila, Dayenne, Thayse e Winnie, por terem me abrigado em suas casas e em seus corações durante este período. Cada dura etapa desta jornada foi suavizada pelo convívio e pela amizade de vocês. Obrigada por tudo!

A todos que compõem a família PPGCTA e que contribuem para o funcionamento do programa.

À equipe do NUPEA-UEPB, em especial a Eliane, Isanna, Flávia, Aline, Michely e Gabriel, por sempre estarem dispostos a ajudar em todos os aspectos. A amizade de vocês tem um valor inestimável!;

A todos os amigos do LAQA, em especial a David, pela amizade, paciência, apoio e por todas as horas de compartilhamento de conhecimento.

Aos produtores de leite caprino e bovino, em especial a CAPRIBOV e a Fazenda Carnaúba, pelo fornecimento das amostras e pelo apoio ao projeto.

Enfim, a todos que contribuíram direta ou indiretamente para a conclusão de uma etapa tão importante da minha vida.

**Um passarinho quando aprende a voar,
sabe mais de coragem que de voo.**

Lucão

RESUMO

A adulteração de leite caprino por adição de leite bovino tem sido uma prática fraudulenta comum que pode acarretar prejuízos econômicos e possíveis danos à saúde do consumidor. Assim, objetivou-se desenvolver metodologias analíticas verdes baseadas em Espectroscopia do Infravermelho Próximo (NIRS) e Imagens Digitais para identificação e/ou quantificação deste tipo de adulteração. Para isto, empregou-se inicialmente um espectrofotômetro NIR convencional de bancada, que, quando acoplado ao Algoritmo das Projeções Sucessivas para seleção de intervalos em regressão por Mínimos Quadrados Parciais (*i*SPA-PLS), foi capaz de quantificar satisfatoriamente este tipo de adulteração, além do teor de lipídeos, enquanto PLS demonstrou os melhores resultados para teor de proteínas. Os resultados de predição apresentaram-se adequados, tendo em vista os elevados coeficientes de correlação e os baixos valores de RMSEP (raiz do erro quadrático médio da predição) e REP (erro relativo da predição), com RPD (relação de desempenho do desvio) maior que 3. Contudo, considerando que, para caracterizar fraude, este tipo de adulteração independe da quantidade de leite bovino adicionada, foram desenvolvidas três diferentes metodologias para sua rápida identificação. Na primeira, utilizando o mesmo equipamento de bancada e PLS para Análise Discriminante (PLS-DA), foi possível identificar adições de leite bovino maiores que 1% (m m^{-1}), classificando corretamente todas as amostras puras e adulteradas em suas respectivas classes. Em seguida, visando portabilizar e simplificar esta metodologia, foi verificada a adequabilidade da utilização de equipamento NIR portátil e de Imagens Digitais como ferramentas analíticas alternativas e de baixo custo para detecção *in situ* deste tipo de adulteração. No caso do NIR portátil, a melhor capacidade preditiva foi observada para *i*SPA-PLS-DA, classificando corretamente 100% das amostras de leite caprino puras e incluindo apenas 1 amostra adulterada na classe de amostras puras. No que diz respeito ao emprego de Imagens Digitais, o melhor desempenho de classificação foi obtido empregando-se o histograma de cor RGB e *Data Driven-Soft Independent Modeling of Class Analogy* (DD-SIMCA), alcançando 90% de taxa de eficiência na discriminação das amostras. Desta forma, ponderando o custo, a acessibilidade e a capacidade discriminante das metodologias propostas, tais técnicas demonstraram-se adequadas como ferramentas rápidas e não destrutivas para triagem da autenticidade de amostras de leite caprino, com aplicabilidade potencial em cooperativas e pequenas indústrias.

Palavras-chave: Algoritmo das Projeções Sucessivas; Autenticidade; Produtos lácteos; NIR portátil; Calibração Multivariada; Reconhecimento de Padrões

ABSTRACT

Adulteration of goat milk by adding cow milk has been a common fraudulent practice that can result in economic losses and possible damage to consumer health. Thus, this work aimed to develop green analytical methodologies based on Near-Infrared Spectroscopy (NIRS) and Digital Images for the identification and/or quantification of this type of adulteration. For this, a conventional benchtop NIR spectrophotometer was used, which was able to satisfactorily quantify both the adulteration and the fat content when coupled with the Successive Projection Algorithm for selection of intervals in Partial Least Squares Regression (*i*SPA-PLS), while PLS demonstrated the best results for the protein content. The prediction results were suitable for this purpose, since they achieved high correlation coefficients and low values of RMSEP (root mean square error of prediction) and REP (relative error of prediction), with RPD (ratio performance deviation) values higher than 3. However, considering that this type of adulteration does not depend on the amount of cow milk added to be characterized as a fraud, three different methodologies were developed for its rapid identification. In the first, using the same benchtop NIR equipment and PLS for Discriminant Analysis (PLS-DA), it was possible to identify additions of cow milk higher than 1% (m m^{-1}), correctly classifying all pure and adulterated samples into their respective classes. Then, in order to make this methodology simple and portable, the suitability of using a miniaturized NIR and Digital Imaging as alternative low-cost analytical tools for the *in situ* detection of this type of adulteration was verified. In the case of the miniaturized NIR, the *i*SPA-PLS-DA attained the best predictive ability, correctly classifying a 100% of the pure goat milk samples and misclassifying only 1 adulterated sample as a pure one. Regarding the use of Digital Images, the best classification performance was obtained using the RGB color histogram and Data Driven-Soft Independent Modeling of Class Analogy (DD-SIMCA), reaching 90% of efficiency for discriminating the samples. Thus, considering both the cost, accessibility, and discriminant ability of the proposed methodologies, such techniques have proven to be adequate as quick and non-destructive tools for screening the authenticity of goat milk samples, with the potential to be applied in cooperatives and small industries.

Keywords: Successive Projections Algorithm; Food authenticity; Dairy Products; Portable NIR; Multivariate Calibration; Pattern Recognition

LISTA DE ILUSTRAÇÕES

ARTIGO 1

Figure 1. NIR spectra from goat (in blue) and cow (in red) milks and their mixtures (in green): (a) Raw and (b) preprocessed spectra with 13 point moving mean coupled with: SNV, (c) MSC, (d) linear baseline correction, (e) linear baseline correction plus baseline offset and (f) baseline offset. 85

Figure 2. (a) Scores and (b) response plots obtained using the NIR spectra pre-processed with 13MM-BO using PCA and PLS-DA, respectively. (c) Contents and (d) biplot of the fat and protein values obtained by the reference methods for the non-adulterated goat (in blue) and non-adulterated cow milk (in red) samples and their mixtures (in green), respectively. 88

Figure 3. (a) Selected intervals and (b) predicted vs reference plot for both calibration (○) and prediction (□) samples (in cyano) using the 20-*i*SPA-PLS/13MM-BO model. (c) Confidence ellipse of the prediction model including the theoretical ideal point. The adulterated samples (mixtures) and non-adulterated goat and cow milk samples are shown in green, blue and red circles, respectively. 92

Figure 4. Selected intervals by 20-*i*SPA-PLS/13MM-LBC for construction of the calibration model for fat content determination. 96

Figure 5. Predicted vs reference plots for both calibration (○) and prediction (□) samples using (a) the 20-*i*SPA-PLS/13MM-LBC model for fat determination and (b) the PLS/13MM-BO-LBC model for protein content, respectively. Confidence ellipses of the prediction models including the theoretical ideal point for the (c) fat and (d) protein content determinations, respectively. 98

Figure S1. (Supplementary Material). (a) Cumulative explained variance and (b) loadings (for five first PCs) plots obtained by PCA using NIR spectra pre-processed with 13MM-BO. 108

ARTIGO 2

Figure 1. NIR spectra from goat (in blue) and cow (in red) milks and their mixtures (in green): raw (a) and pre-processed spectra with 7-point moving mean coupled with: standard normal variation (b), multiplicative scattering correction (c), linear baseline correction (d), linear baseline correction plus baseline offset (e) and baseline offset (f).

117

Figure 2. Scores (a) and loadings (b) plots obtained by Principal Component Analysis using the NIR spectra pre-processed with 7-point moving mean and baseline offset correction.

119

Figure 3. Training (a) and prediction (b) acceptance area for the OC-PLS model using the NIR spectra pre-processed with smoothing by 7-point moving mean and baseline offset correction. Plots of the error rate as a function of the latent variables (c) and its response (d) for the PLS-DA model using the NIR spectra pre-processed with smoothing by 7-point moving mean and multiplicative scattering correction. Plots of the intervals selected by SPA (e) and its response (f) for the 10-*i*SPA-PLS-DA model using the NIR spectra pre-processed with smoothing by 7-point moving mean and baseline offset correction. Orange lines indicate the borders of the area of acceptance in OC-PLS and the thresholds between the studied classes in PLS-DA and 10-*i*SPA-PLS-DA. The pure goat milk samples are represented in blue and the adulterated samples in green.

121

ARTIGO 3

Figura 1. (a) Ilustrações da mesma amostra de leite caprino sem e com a adição de leite bovino. (b) Histogramas de cor médios obtidos para as amostras puras (linha azul) e adulteradas (linha verde) de leite caprino nos diferentes canais analíticos: R – vermelho, G – verde, B – azul, H – matiz, S – saturação, I – intensidade.

138

Figura 2. Análise de Componentes Principais: gráficos (a, b, c, d) de escores dos histogramas de cor GS, RGB, HSI e GS+RGB+HSI, respectivamente e (e, f, g, h) pesos (loadings) obtidos para as duas primeiras componentes principais (PC1 – linha preta e PC2 – linha roxa). As amostras puras de leite caprino são representadas em azul e as adulteradas com leite bovino em verde. 140

Figura 3. Gráficos de área de aceitação para o conjunto de (a,c) treinamento e (b,d) teste empregando OC-PLS para o histograma GS e DD-SIMCA para histograma RGB, respectivamente. As bolas azuis e verdes representam amostras de leite caprino puras e adulteradas, respectivamente. As linhas laranjas representam a borda entre as amostras regulares e extremas; a linha púrpura é a fronteira da área de aceitação. 144

Figura S1 – Gráficos das variâncias explicadas em função das componentes principais, para os histogramas de cor GS (escala de cinza), RGB (vermelho, verde e azul), HSI (matiz, saturação e intensidade) individuais e combinados (GS+RGB+HSI). 155

LISTA DE TABELAS

ARTIGO 1

Table 1. Description of the studied goat and cow milk samples.	81
Table 2. Results of the determinations of cow milk in goat milk using different pre-processed NIR spectra and PLS algorithms.	90
Table 3. Results of the determinations of fat content in cow milk, goat milk and its mixtures using different pre-processed NIR spectra and PLS algorithms.	94
Table 4. Results of the determinations of protein content in cow milk, goat milk and its mixtures using different pre-processed NIR spectra and PLS algorithms.	95

ARTIGO 2

Table 1. Description of the studied goat and cow milk samples.	114
Table 2 - Classification results of the goat milk adulteration with cow milk using different pre-processed NIR spectra and PLS-based algorithms.	120

ARTIGO 3

Tabela 1. Descrição das amostras de leite caprino e bovino analisadas.	135
Tabela 2. Resultados de sensibilidade, especificidade e eficiência para OC-PLS e DD-SIMCA nos conjuntos de treinamento e teste.	142

LISTA DE QUADROS

Quadro 1- Metodologias analíticas utilizadas para determinação da autenticidade de leite caprino e derivados.	30
Quadro 2- Descrição das amostras utilizadas na identificação e quantificação de adulteração de leite caprino por adição de leite bovino, por espectroscopia NIR e imagens digitais.	46

LISTA DE ABREVIACES

BO	Correo <i>offset</i> de linha de base (do ingls, <i>Baseline Offset</i>)
CLA	cido linoleico conjugado (do ingls, <i>Conjugated linoleic acid</i>)
CN	Casena
DD-SIMCA	<i>Data Driven-Soft Independent Modeling of Class Analogy</i>
DPI	Pontos por polegada (do ingls, <i>Dots per inch</i>)
EJCR	Regio de confiana da elipse (do ingls, <i>Elliptical joint confidence region</i>)
GS	Escala de cinza (do ingls, <i>Grayscale</i>)
HSB	Matiz, saturao, brilho (do ingls, <i>Hue, saturation, brightness</i>)
HSI	Matiz, saturao, intensidade (do ingls, <i>Hue, saturation, intensity</i>)
HSL	Matiz, saturao, luminosidade (do ingls, <i>Hue, saturation, lightness</i>)
iPLS	PLS com seleo de intervalo (do ingls, <i>Interval PLS</i>)
iSPA-PLS	PLS combinado com algoritmo de sucessivas projees para seleo de intervalos (do ingls, <i>Successive projections algorithm - PLS</i>)
KS	Kennard-Stone
LBC	Correo linear de linha de base (do ingls, <i>Linear baseline correction</i>)
MM	Mdia mvel (do ingls, <i>Moving mean</i>)
MSC	Correo do espalhamento multiplicativo (do ingls, <i>Multiplicative scatter correction</i>)
NIR	Infravermelho prximo (do ingls, <i>Near Infrared</i>)
NIRS	Espectroscopia no infravermelho prximo (do ingls, <i>Near Infrared Spectroscopy</i>)
NUPEA	Ncleo de Pesquisa e Extenso em Alimentos
OC-PLS	<i>One-class-PLS</i>
PCA	Anlise de componentes principais (do ingls, <i>Principal component analysis</i>)

PCR	Reação em cadeia da polimerase (do inglês, <i>Polymerase chain reaction</i>)
PLS	Regressão dos mínimos quadrados parciais (do inglês, <i>Partial Least Squares</i>)
PLS-DA	Análise discriminante PLS (do inglês, <i>PLS- Discriminant analysis</i>)
r_{cv}	Coeficiente de correlação da validação cruzada
REP	Erro relativo da predição (do inglês, <i>Relative error of prediction</i>)
RGB	Vermelho, verde e azul (do inglês, <i>Relative error of prediction</i>)
RMSECV	Raiz do erro quadrático médio da validação cruzada (do inglês, <i>Root mean square error of crossvalidation</i>)
RMSEP	Raiz do erro quadrático médio da predição (do inglês, <i>Root mean square error of prediction</i>)
RPD	Relação de desempenho do desvio (do inglês, <i>Ratio performance deviation</i>)
r_{pred}	Coeficiente de correlação da predição
SNV	Variação normal padrão (do inglês, <i>Standard normal variate</i>)
TCM	Triglicerídeos de cadeia média
VL	Variáveis latentes

SUMÁRIO

1 INTRODUÇÃO	20
2 REVISÃO DE LITERATURA	23
2.1 CAPRINOCULTURA LEITEIRA E SUAS POTENCIALIDADES	23
2.2 LEITE CAPRINO	24
2.2.1 Fração proteica	25
2.2.2 Fração lipídica	27
2.2.3 Adulterações	28
2.2.4 Metodologias analíticas para identificação de autenticidade de leite caprino	29
2.3 ESPECTROSCOPIA DE INFRAVERMELHO PRÓXIMO (NIRS)	31
2.3.1 Fundamentos da técnica	31
2.3.2 Características físicas das amostras	33
2.3.3 Variabilidade química das amostras	37
2.3.4 Análise de dados (Quimiometria)	39
2.4 IMAGENS DIGITAIS	43
3 MATERIAIS E MÉTODOS	46
3.1 AMOSTRAS	46
3.2 EXPERIMENTOS	47
3.2.1 Experimento 1- Equipamento NIR de bancada	47
3.2.2 Experimentos 2 (Equipamento NIR portátil) e 3 (Imagens Digitais)	48
3.3 OBTENÇÃO DAS AMOSTRAS ADULTERADAS DE LEITE CAPRINO	49
3.4 OBTENÇÃO DOS ESPECTROS NIR	49
3.5 AQUISIÇÃO DAS IMAGENS DIGITAIS	50
3.6 ANÁLISE QUIMIOMÉTRICA DOS DADOS	50
3.6.1 Tratamento dos dados - Experimento 1	51
3.6.2 Tratamento dos dados – Experimento 2	52
3.6.3 Tratamento dos dados – Experimento 3	53
3.7 ANÁLISES FÍSICO-QUÍMICAS DAS AMOSTRAS DE LEITES CAPRINO E BOVINO	54
REFERÊNCIAS	55
4 RESULTADOS	75
ARTIGO I: Simultaneous determination of goat milk adulteration with cow milk and their fat and protein contents using NIR spectroscopy and PLS algorithms	75
ARTIGO II: A low-cost portable NIR spectrophotometer for <i>in situ</i> identification of goat milk adulteration with cow milk	109
ARTIGO III Imagens digitais como ferramenta para autenticação de leite caprino em termos de adulteração por adição de leite bovino: um estudo de viabilidade	130
5 CONSIDERAÇÕES FINAIS	156

1 INTRODUÇÃO

A produção mundial de leite caprino ultrapassou, em 2017, 18 milhões de toneladas por ano, acumulando um aumento de aproximadamente 20% em uma década (2007-2017) (FAOSTAT, 2019). Esse aumento da produção se deve, em parte, ao crescente interesse do consumidor por este produto e seus derivados, motivado pelas características singulares do leite caprino quando comparado ao leite bovino, tais como baixa alergenicidade, alta digestibilidade e propriedades funcionais (AHMED et al., 2015; CLARK; GARCÍA, 2017; VERRUCK; DANTAS; PRUDENCIO, 2019).

Para atender esta demanda, estudos recentes têm focado no desenvolvimento de novos produtos a base de leite de cabra, dentre os quais se destacam iogurtes, queijos, sorvetes e doces (MCGHEE; JONES; PARK, 2015; SILVA et al., 2015; SERHAN; MATTAR; DEBS, 2016; SATIR; GUZEL-SEYDIM, 2016). Contudo, do ponto de vista econômico, o leite de cabra é mais caro que o leite de vaca e, consequentemente, seus derivados possuem um maior valor agregado. Desta forma, sua representatividade econômica contribui para que as práticas de adulteração de leite caprino sejam uma realidade (RODRIGUES et al., 2012; GOLINELLI et al., 2014; AGRIMONTI et al., 2015; DI PINTO et al., 2017).

A maior parte das fraudes se enquadra na categoria de substituição, na qual há uma substituição parcial ou completa de um constituinte do alimento por um composto de menor custo sem o conhecimento dos consumidores, como acontece, por exemplo, com a adição de um adulterante ou mistura de adulterantes (MOORE; SPINK; LIPP, 2012). No caso do leite caprino e seus derivados, o principal adulterante encontrado é o leite bovino, devido ao seu menor custo e maior abundância (DABROWSKA et al., 2010; RODRIGUES et al., 2012; GOLINELLI et al., 2014; CHOOPAN; THANAKIATKRAI; KITPIPIT, 2017; DI PINTO et al., 2017; KIM; KIM, 2018). Além disso, considerando a matriz alimentícia leite no âmbito geral, uma gama de adulterantes também tem sido relatada, tais como soro de leite, melamina, amido, sulfato de amônio e ureia (BILGE et al., 2016; LIM et al., 2016; NASCIMENTO et al., 2017; TRIPATHY et al., 2017).

Apesar de bem descrita e relatada desde a década de 80, a preocupação com a determinação de adulterantes em leite ainda é um tema atual. Esta tendência é comprovada pelo crescente número de publicações acadêmicas sobre este tema em

periódicos de alto fator de impacto no período de 2010 a 2017 (NASCIMENTO et al., 2017; DI PINTO et al., 2017).

Neste contexto, diferentes metodologias têm sido desenvolvidas e validadas com o objetivo de identificar e quantificar adulterações em leite, tendo em vista os potenciais impactos econômicos e sobre a saúde do consumidor, dada a natureza tóxica de alguns contaminantes (MOORE; SPINK; LIPP, 2012; KAMAL; KAROUI, 2015).

A avaliação de parâmetros de qualidade de leite exigidos pelos órgãos de fiscalização para autorizar a comercialização do produto tem sido realizada com base nas suas características físico-químicas e microbiológicas (MAPA, 2000; 2006; 2018a; 2018b). Alternativamente, técnicas cromatográficas têm se tornado os métodos mais comumente utilizados no rastreamento da autenticidade de produtos lácteos (KAMAL; KAROUI, 2015). Contudo, estas técnicas possuem desvantagens, tais como o elevado custo, a lentidão de análise e a necessidade de várias etapas de pré-tratamento das amostras, as quais requerem, muitas vezes, grandes quantidades de reagentes e solventes.

Neste cenário, a Espectroscopia de Infravermelho Próximo (NIRS) tem se destacado por ser uma técnica útil para a detecção de adulteração de diferentes tipos de alimentos, dentre os quais o leite e seus derivados se destacam (LOHUMI et al., 2015; LIM et al., 2016; CHEN et al., 2017). De forma complementar, visto que produtos alimentícios são matrizes complexas e que cada metodologia tem suas vantagens e limitações, a associação de diferentes técnicas analíticas para obtenção de informações mais precisas e reprodutíveis tem sido relatada como uma alternativa promissora (KAMAL; KAROUI, 2015; LOHUMI et al., 2015).

Ainda no campo das técnicas não destrutivas, as Imagens Digitais também têm se tornado uma ferramenta analítica potente, uma vez que podem realizar o controle de qualidade de alimentos, tanto para fins qualitativos quanto quantitativos (ROMANO et al., 2012; EL-SAYED et al., 2016). Não obstante, estas técnicas geram uma grande quantidade de informações por amostra, o que requer o uso de ferramentas quimiométricas multivariadas para extração da informação útil. Ainda assim, tais técnicas demonstram habilidade para detectar adulterações em alimentos com boa precisão, o que as tornam alternativas viáveis para a substituição das metodologias de referência (PASQUINI, 2018; KAMAL; KAROUI, 2015).

Ante o exposto e considerando que, para o tema adulteração em produtos lácteos, os relatos para leite caprino e derivados ainda são uma vertente a ser explorada,

objetivou-se com este trabalho desenvolver metodologias analíticas verdes baseadas em NIRS e imagens digitais para identificação e quantificação de adulteração de leite caprino por adição de leite bovino.

2 REVISÃO DE LITERATURA

2.1 CAPRINOCULTURA LEITEIRA E SUAS POTENCIALIDADES

Atualmente, a Ásia é o continente responsável pela maior parte da produção mundial de leite caprino, representando 56,6% do total, seguido da África, Europa e das Américas. Em termos de países produtores, a Índia se destaca, tendo alcançado produção superior a 6 milhões de toneladas em 2017, o que representa aproximadamente 1/3 da produção mundial. Dentre os países europeus, a França desponta como principal produtor (5º maior produtor mundial), com produção superior a 590 mil toneladas em 2017 (FAOSTAT, 2019).

Esse cenário fica um pouco diferente quando analisamos o mercado produtor de derivados lácteos de origem caprina. No que diz respeito aos queijos de leite caprino, principal derivado em termos de consumo, África e Europa são responsáveis por mais de 75% da produção mundial, tendo o Sudão e a França como principais produtores (FAOSTAT, 2019). Esses dados refletem as dinâmicas de produção e consumo do leite caprino e seus derivados no cenário mundial, demonstrando o potencial econômico a ser explorado pelos grandes produtores mundiais.

No que se refere ao panorama brasileiro, a produção nacional de leite caprino ultrapassou 250 mil toneladas em 2017, o que não corresponde, entretanto, ao pico produtivo do país na última década. Em 2013, foram produzidas mais de 273 mil toneladas de leite caprino, com uma queda significativa superior a 10% da produção no ano seguinte. Desde então, o país vem retomando o crescimento anual dessa produção (FAOSTAT, 2019).

As regiões Nordeste e Sudeste são as maiores produtoras de leite caprino no Brasil, sendo responsáveis por 92% do total produzido. Em termos de estados produtores, a Paraíba se destaca como maior produtor de leite caprino, tendo produção superior a 5 milhões de litros em 2017. Os 5 municípios com maior produção foram Monteiro, Amparo, Taperoá, Sumé e Parari, todos localizados na região do Cariri paraibano (IBGE, 2018).

Vale ressaltar, entretanto, que as regiões diferem substancialmente nas características da cadeia produtiva e na destinação da produção. Enquanto a região

Nordeste, responsável 75% da produção nacional, produz, em larga escala, apenas leite pasteurizado e destina sua produção majoritariamente a programas dos governos Federal e Estadual (Programa de Aquisição de Alimentos – PAA), a região Sudeste, que representa 17% do total, possui uma linha mais diversificada de produtos (leite UHT, leite em pó, iogurtes e queijos), verticalizando sua produção e atendendo a nichos de mercados dos grandes centros urbanos (FELISBERTO; OLIVEIRA; CORDEIRO, 2016; FAOSTAT, 2019).

Neste contexto, torna-se evidente que, assim como no cenário mundial, no Brasil o maior retorno econômico não está associado diretamente aos números de produção, tendo em vista que a industrialização e a produção de derivados de maior valor agregado ainda estão compartimentalizadas no país.

Muitos são os entraves para que a caprinocultura leiteira se torne mais expressiva em termos econômicos no Brasil. Além do aspecto cultural, que dificulta a inserção dos produtos no mercado, a carência de serviços de assistência técnica e extensão rural dificultam a inserção de tecnologias, monitoramento e avaliação da eficiência dos sistemas (FELISBERTO; OLIVEIRA; CORDEIRO, 2016).

Desta forma, o desenvolvimento de tecnologias e metodologias que atinjam esta atividade rural, em todos os seus aspectos, quer seja na viabilização da produção, aproveitamento ou monitoramento dos aspectos de qualidade, tem potencial para o fortalecimento do mercado produtor, com vistas ao desenvolvimento regional.

2.2 LEITE CAPRINO

De acordo com a Instrução Normativa Nº 51/2002 do Ministério da Agricultura Pecuária e Abastecimento (MAPA), entende-se por leite, sem qualquer outra denominação, o produto oriundo da ordenha completa, ininterrupta, em condições de higiene, de vacas sadias, bem alimentadas e descansadas. O leite proveniente de outras espécies deve denominar-se segundo a espécie da qual proceda. Por sua vez, para a denominação leite caprino temos a seguinte definição: produto oriundo da ordenha completa, ininterrupta, em condições de higiene, de animais da espécie caprina sadios, bem alimentados e descansados (MAPA, 2000).

Em média, o leite caprino é composto por 87% de água, 3,6% de proteínas, 4,3% de lipídeos, 4,3% de carboidratos e 0,8% de cinzas (CLAEYS et al., 2014; ALICHANIDIS; MOATSOU; POLYCHRONIADOU, 2016). Entretanto, a composição química do leite de cabra pode variar de acordo com a raça, alimentação, sistema de produção (extensivo e intensivo), período de lactação, condições ambientais, idade e estado de saúde dos animais (STEINSHAMN et al., 2014; TABET et al., 2016; SATIR; GUZEL-SEYDIM, 2016; CAROPRESE et al., 2016; MCLEOD; PROSSER; WAKEFIELD, 2016; EL-TARABANY; EL-TARABANY; ROUSHDY, 2016).

Considerando os constituintes de maior abundância no leite caprino (lactose, proteínas e lipídeos), as diferenças mais relevantes em comparação com o leite bovino, como esperado, são relacionadas aos perfis proteico e lipídico, tendo em vista a complexidade e variabilidade potencial dessas substâncias e, por este motivo, estas frações serão detalhadas nos itens subsequentes.

2.2.1 Fração proteica

Na fração proteica do leite é possível destacar, em termos de abundância, duas classes: a caseínica e a das proteínas do soro. Em cabras e vacas, bem como em outros ruminantes, a proporção média das classes no teor de proteína dos leites é superior a 80% para a fração caseínica e da ordem de 15% para as proteínas do soro (CLAEYS et al., 2014; ALICHANIDIS; MOATSOU; POLYCHRONIADOU, 2016). As características gerais de ambas as frações proteicas do leite encontram-se bem descritas na literatura (HOLT, 2016; DAVE; SINGH, 2019; DEETH; BANSAL, 2019). Neste tópico, abordar-se-á apenas as informações relevantes para diferenciação dos leites caprino e bovino, tendo em vista a relevância para o tema estudado.

Considerando a fração caseínica, enquanto componente majoritário da porção proteica, os leites caprino e bovino podem apresentar variações intra e interespecíficas, tanto sob o aspecto quantitativo quanto qualitativo (EGITO et al., 2006; CEBALLOS et al., 2009; COSTA et al., 2014; PERINELLI et al., 2019). Esta variação está relacionada ao elevado índice de polimorfismo no DNA dos 4 genes da família da caseína, a citar CSN1S1, CSN2, CSN1S2 e CSN3 (GÓMEZ-RUIZ et al., 2004; MONTALBANO et al., 2016; WANG et al., 2018b). Em cabras, por exemplo, já foram descritas pelo menos 18

variantes para o gene que codifica a fração α_{s1} (CSN1S1) e 7 para o gene que codifica a fração α_{s2} (CSN1S2), incluindo alelos nulos, os quais resultam na ausência destas frações no leite (GÓMEZ-RUIZ et al., 2004; TOMATOKE et al., 2006; ALBENZIO et al., 2012; CEBO et al., 2012; COSTA et al., 2014; VERRUCK; DANTAS; PRUDENCIO, 2019).

No que diz respeito à similaridade das proteínas, a homologia da sequência de aminoácidos das frações das caseínas dos ruminantes é, em geral, superior a 85%, podendo alcançar, para algumas frações, índices superiores a 96%, apresentando, entretanto, diferentes níveis de fosforilação (GINGER; GRIGOR, 1999; PERINELLI et al., 2019). Esta similaridade pode ser evidenciada tanto em perfis cromatográficos e eletroforéticos obtidos a partir dos leites bovino e caprino, quanto na característica reativa a anticorpos das proteínas de ambas as origens, salvaguardadas as devidas peculiaridades de cada espécie (EGITO et al., 2006; RODRÍGUEZ et al., 2010; PESIC et al., 2011; EL-HATMI et al., 2015; PERINELLI et al., 2019).

Neste contexto, é importante destacar algumas das características singulares das espécies em questão, pois estas são frequentemente utilizadas como marcadores potenciais na determinação de adulteração por misturas de leites através de diferentes metodologias.

Chen et al. (2004) e Pesic et al. (2011) identificaram 2 variantes da proteína β -lactoglobulina (A e B) no leite bovino, enquanto no leite caprino apenas 1 tipo foi verificado. Além disso, estes autores também relataram diferenças entre as espécies no que diz respeito às massas e à polaridade/mobilidade de α -lactoalbumina e β -lactoglobulina. De forma análoga, Molina, Martín-Álvarez e Ramos (1999) verificaram diferenças entre os leites caprino e bovino quanto aos tempos de migração das frações caseínicas α_{s2} -CN, α_{s1} -CN, κ -CN e β -CN.

No tocante ao perfil de aminoácidos, Ceballos et al. (2009) e Rafiq et al. (2016) relataram diferenças significativas entre amostras de leite caprino e bovino, no que concerne os teores dos aminoácidos identificados. Apesar da similaridade do tema estudado, estes autores apresentam resultados discrepantes quanto à correlação entre a espécie e o teor de determinado aminoácido, fato que pode estar associado às características das amostras estudadas relacionadas aos fatores que podem influenciar a composição do leite.

Além disso, os diferentes pontos isoelétricos das frações γ da caseína, originadas a partir da hidrólise enzimática da β -caseína, nas espécies caprina e bovina também são

utilizados como ferramenta de diferenciação do leite dessas espécies. Esta, inclusive, é a característica na qual o princípio do método oficial estabelecido pela União Europeia para detecção de leite bovino em queijos caprinos e ovinos é baseado (European Commission 2018/150, 2018).

2.2.2 Fração lipídica

Tendo em vista que o conteúdo lipídico é o componente orgânico mais variável da matéria seca do leite das diferentes espécies (CLAEYS et al., 2014; ALICHANIDIS; MOATSOU; POLYCHRONIADOU, 2016), diversos estudos estabelecem correlações entre o teor e/ou perfil desta fração com os diferentes fatores que afetam a composição do leite (CHEN; LEWIS; GRANDISON, 2014; LI et al., 2017; KHOLIF; MORSY; ABDO, 2018; SANT'ANA et al., 2019; YURCHENKO et al., 2018; ALMEIDA et al., 2019), de forma que este componente é um notável contribuinte para a variação intra e interespecífica do leite das diferentes espécies.

É importante destacar, entretanto, que as mudanças na secreção de lipídeos no leite, relacionadas a fatores nutricionais e fisiológicos, diferem entre caprinos e bovinos. Chilliard et al. (2014), através do levantamento bibliográfico de publicações sobre este tema, relataram que, enquanto o estágio de lactação tem efeitos equivalentes sobre o teor de lipídeos em cabras e vacas, a influência da dieta pode diferir entre as espécies, de tal forma que, em termos gerais, a composição lipídica do leite caprino é menos influenciada pela dieta que a do leite bovino. Além disso, apesar da dependência do teor e da composição dos suplementos lipídicos implementados na dieta de ambas espécies, os efeitos e os mecanismos são diferentes, resultando em diferenças na secreção e no perfil de lipídeos.

No âmbito geral, dentre as diferenças mais relatadas nos estudos que comparam os perfis lipídicos do leite caprino com o bovino, é possível citar: maiores níveis de ácidos graxos insaturados (monoinsaturados + poli-insaturados), destacando-se os teores de ω -3 e ω -6; maior teor de ácido linoleico conjugado (CLA); maior conteúdo de ácidos graxos de cadeia média (TCM) e curta; e maior razão entre os TCM e o conteúdo total de triglicerídeos (TOMATOKE et al., 2006; CEBALLOS et al., 2009; SERHAN;

MATTAR; DEBS, 2016; VIEITEZ et al., 2016; LI et al., 2017). Tais características estão frequentemente associadas aos benefícios do leite caprino em relação ao leite bovino.

Além das características químicas do perfil lipídico, diferentes autores relatam que o tamanho dos glóbulos de gordura do leite caprino é, em média, menor que os do leite bovino. Considerando que, quanto menor o tamanho do glóbulo, mais eficiente é o metabolismo lipídico, tal característica está associada a maior digestibilidade do leite caprino (CLAEYS et al., 2014; ALICHANIDIS; MOATSOU; POLYCHRONIADOU, 2016).

Neste contexto, diversos autores sugerem que tais diferenças podem ser utilizadas como marcadores potenciais da presença de adulteração por misturas de leites (CEBALLOS et al., 2009; VIEITEZ et al., 2016; GONZÁLEZ-MARTÍN et al., 2017; LI et al., 2017). Vale ressaltar, entretanto, que, assim como para proteínas, dada a variância intra e interespecífica dos teores e da composição lipídica, é necessário considerar a representatividade amostral dos estudos que se propõem a identificar tais adulterações utilizando apenas esta fração da matriz alimentícia.

2.2.3 Adulterações

A adulteração de leite geralmente envolve diluição e/ou adição de ingredientes mais baratos e de baixa qualidade e, em alguns casos, que causam risco à saúde humana, com o objetivo de aumentar o volume, mascarar qualidade inferior ou substituir substâncias naturais do leite para ganho econômico (NASCIMENTO et al., 2017).

Segundo a legislação brasileira (MAPA, 2000), será considerado fraudado ou falsificado o leite caprino que tiver sofrido: (i) a adição, subtração ou substituição de quaisquer componentes naturais ou estranhos à composição normal do leite de cabra, sem a devida aprovação prévia do Serviço de Inspeção Federal e declaração na rotulagem; (ii) adição ou presença de substâncias conservadoras ou inibidoras do crescimento de microrganismos; (iii) adição de leite de outras espécies animais.

Devido a maior abundância e ao menor custo, a adição de leite bovino é a adulteração mais comumente relatada em leite caprino e seus derivados (DABROWSKA et al., 2010; RODRIGUES et al., 2012; GOLINELLI et al., 2014;

AGRIMONTI et al., 2015; DI PINTO et al., 2017). Por este motivo, para fins de exemplificação e caracterização deste problema, relatamos aqui alguns dos resultados das pesquisas sobre este tipo de adulteração.

Dabrowska et al. (2010) detectaram a presença de leite bovino em leite caprino em 77% das amostras analisadas, com concentrações variando entre 1% a 20% em estudo realizado na Polônia. Rodrigues et al. (2012) determinaram a presença de leite bovino em 41,2% das amostras de leite caprino provenientes do Nordeste do Brasil. Golinelli et al. (2014) avaliaram 20 amostras de queijos de cabra disponíveis comercialmente no Rio de Janeiro (Brasil) e detectaram a presença de leite bovino em todas as amostras analisadas. Agrimonti et al. (2015) verificaram a presença de leite bovino em amostras de iogurte, queijo e manteiga caprina comercializadas na Itália. Di Pinto et al. (2017) identificaram adulteração de leite e queijos caprinos italianos por adição de leite bovino em 80% das amostras avaliadas.

Vale ressaltar, ainda, que a matriz leite, avaliada em toda sua amplitude, é um dos alimentos mais fraudados, com os mais diversos tipos de aditivos (MOORE; SPINK; LIPP, 2012). Levantamentos bibliográficos recentes demonstram que adulteração do leite, apesar de bem descrita e relatada ao longo dos anos, ainda é um tema atual e de grande relevância (HANDFORD; CAMPBELL; ELLIOTT, 2016). Dentre os adulterantes sintéticos, os mais comumente encontrados são ureia, açúcar, melamina, formaldeído, peróxido de hidrogênio e detergentes (SOUZA et al., 2011; NIRWAL; PANT; RAI, 2013; SCHODER, 2010; HANDFORD; CAMPBELL; ELLIOTT, 2016).

2.2.4 Metodologias analíticas para identificação de autenticidade de leite caprino

A determinação da autenticidade e a detecção de adulteração de leite e produtos lácteos têm sido realizadas por métodos analíticos convencionais, tais como físico-químicos, sensoriais e cromatográficos (FINETE et al., 2013; KAMAL; KAROUI, 2015; VIEITEZ et al., 2016). Além destes, metodologias que utilizam as ferramentas da biologia molecular também têm sido consideradas promissoras para identificação e quantificação de adulterações em leite e derivados. No que diz respeito aos métodos baseados em amplificação de DNA por PCR (do inglês, *Polymerase Chain Reaction*), é

possível citar os estudos de López-Calleja et al. (2004), Dabrowska et al. (2010), Rodrigues et al. (2012) e Di Pinto et al. (2017), relativos à quantificação no primeiro estudo e à identificação nos demais trabalhos de adulteração de leite caprino por adição de leite bovino, através da amplificação de regiões bem conservadas do DNA mitocondrial.

Além disso, outros autores também relataram a adequabilidade de técnicas eletroforéticas (CARTONI et al., 1999; PESIC et al., 2011) e de ensaios imunológicos (SONG; XUE; HAN; 2011) para esta finalidade, apresentando elevados coeficientes de correlação entre os resultados das análises e a concentração de leite bovino presente no leite caprino.

Entretanto, diante da relevância do tema adulteração da matriz alimentícia leite e das limitações das técnicas convencionais (e.g. alto custo, elevada demanda de reagentes e necessidade de operadores capacitados), são inúmeros os esforços na vertente de desenvolvimento de métodos analíticos mais rápidos e baratos capazes de detectar, identificar e quantificar tais adulterantes (KAMAL; KAROUI, 2015; HANDFORD; CAMPBELL; ELLIOTT, 2016).

No contexto das metodologias não destrutivas, as técnicas espectroscópicas, tais como Espectroscopias de Fluorescência, Infravermelho Próximo e Médio (NIR e MIR, (do inglês, *Near Infrared and Mid-Infrared Spectroscopy*) respectivamente), Ressonância Magnética Nuclear (NMR, do inglês, *Nuclear Magnetic Resonance*) e Imagens Digitais em combinação com a Quimiometria e, alternativamente com outras técnicas, têm sido alternativas amplamente exploradas na área de determinação de adulterantes em leite (DIAS et al., 2009; SANTOS; PEREIRA-FILHO, 2013; KAMAL; KAROUI, 2015; LOHUMI et al., 2015; BRANDÃO; ANJOS; BELL, 2017; CHEN et al., 2017; GENIS et al., 2020; ULLAH et al., 2020). O **Quadro 1** apresenta um resumo das diferentes metodologias analíticas empregadas na determinação da autenticidade de leite caprino e derivados.

Quadro 1- Metodologias analíticas utilizadas para determinação da autenticidade de leite caprino e derivados.

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

2.3 ESPECTROSCOPIA DE INFRAVERMELHO PRÓXIMO (NIRS)

2.3.1 Fundamentos da técnica

A Espectroscopia de Infravermelho Próximo (NIRS) é um tipo de espectroscopia vibracional que emprega um feixe eletromagnético na faixa de comprimentos de onda entre 750 a 2500 nm (13300 a 4000 cm^{-1}) e mede a energia absorvida pelas moléculas. O objetivo geral é adquirir informações qualitativas e quantitativas das amostras a partir da interação entre seus constituintes e as ondas eletromagnéticas na região do infravermelho próximo (PASQUINI, 2003).

Os métodos analíticos resultantes do uso da NIRS possuem características relevantes por serem rápidos, não destrutivos, não invasivos, com alta penetração da

radiação, aplicabilidade quase universal (qualquer molécula contendo ligações C-H, N-H, S-H e/ou O-H), com uma demanda mínima de preparação de amostras, o que os tornam vantajosos e viáveis para aplicação industrial *on-line* (PASQUINI, 2018; POREP; KAMMERER; CARLE, 2015).

As características, aplicações, bem como vantagens e limitações desta técnica estão bem descritas e consolidadas em importantes revisões (PASQUINI, 2003; POREP; KAMMERER; CARLE, 2015; PASQUINI, 2018). Não obstante todas as vantagens metodológicas já mencionadas anteriormente, assim como a maioria das análises experimentais, a NIRS possui suas limitações, principalmente no que diz respeito à sensibilidade da técnica (PASQUINI, 2018) e dependência de certos fatores operacionais e ambientais, tais como temperatura, variabilidade e estado físico da amostra (ELY; THOMMES; CARVAJAL, 2008; POJIC; MASTILOVIC; MAJCEN, 2012; DUPONT; CROGUENNEC; POCHET, 2018).

Como a NIRS utiliza radiação eletromagnética de baixa intensidade e, conseqüentemente, gera respostas de baixa intensidade das moléculas excitadas, a sensibilidade dos métodos baseados nesta metodologia precisa ser considerada, pois a torna inadequada para análises de substâncias traço com concentração abaixo de 1% (KAROUI; DE BAERDMAEKER, 2007; WANG et al, 2017; MANLEY; BAETEN, 2018; PASQUINI, 2018).

Neste contexto analítico, estabelecer elevadas taxas entre o sinal e o ruído é um desafio para diferenciação entre os espectros das amostras (KAROUI; DE BAERDMAEKER, 2007; MANLEY; BAETEN, 2018). O ruído pode ser causado pela variação do equipamento durante a análise, pela preparação e apresentação da amostra, bem como pelas condições ambientais dentro das quais os espectros são gerados (MANLEY; BAETEN, 2018). Desta forma, buscar a redução do ruído, quer seja pela utilização de ferramentas quimiométricas de seleção de variáveis ou pela padronização das amostras, tem sido determinante nos estudos com aplicação da espectroscopia NIR em alimentos.

De forma geral, estudos relacionados com a adulteração de leite e derivados limitam-se a determinar substâncias adulterantes em concentrações acima do limite de quantificação adequado para a técnica (MABOOD et al., 2017a; MABOOD et al., 2017b). Entretanto, alguns autores já relataram a construção de modelos preditivos para concentrações de adulterantes abaixo de 1%, através da combinação da NIRS convencional com Quimiometria, utilizando diferentes métodos de regressão linear

(WU et al., 2016; CHEN et al., 2017; KARUNATHILAKA et al., 2018), e na associação da espectroscopia NIR com Imageamento, como no uso de NIR Hiperespectral (FU et al., 2014; HUANG et al., 2016; LIM et al., 2016; FORCHETTI e POPPI, 2017).

Vale ressaltar, entretanto, que os estudos que relataram a construção de modelos preditivos na classificação de amostras adulteradas de leite, em concentrações abaixo de 1%, através do uso do NIR associado com ferramentas quimiométricas, referem-se à identificação de adulterantes sintéticos, que não estão naturalmente presentes na matriz láctea, tais como melamina, biureto e aminotriazol (WU et al., 2016; CHEN et al., 2017; KARUNATHILAKA et al., 2018). Complementarmente, Karunathilaka et al. (2018) demonstraram que, quando são realizadas misturas de leite com adulterantes sintéticos, é possível identificar visualmente diferenças significativas nos espectros NIR entre amostras adulteradas e puras, diferentemente do que acontece quando são empregados adulterantes naturais (mesmo que proveniente de outras origens), tais como proteína de soja ou ervilha, quando os espectros NIR são mais similares. Este é, provavelmente, um dos motivos pelo qual estudos nos quais o adulterante é um leite de outra espécie, onde espera-se um nível de similaridade elevado, as concentrações testadas sejam sempre superiores ao limite de detecção da técnica (MABOOD et al., 2017a; MABOOD et al., 2017b).

Além disso, na maioria dos casos, ferramentas quimiométricas de seleção de variáveis para extração da informação relevante e redução do ruído demonstraram-se imprescindíveis para a obtenção de modelos mais ajustados para concentrações abaixo de 1% (WU et al., 2016; CHEN et al., 2017).

2.3.2 Características físicas das amostras

Como citado anteriormente, alguns dos fatores que afetam a NIRS estão relacionados com as características físicas de apresentação das amostras e com algumas condições operacionais adotadas. Desta forma, abordaremos aqui os principais efeitos dessas características, bem como sua abordagem nos estudos mais recentes sobre adulteração de leite e derivados baseados na NIRS.

Considerando a variedade e a composição complexa dos produtos lácteos, é possível ponderar a influência de alguns desses fatores na construção de métodos para determinação de adulterações baseados na NIRS. Neste contexto, a natureza da amostra e as condições de processamento do produto e da amostra podem ser considerados os fatores mais relevantes.

No tocante ao estado físico da amostra, a apresentação do produto (líquido ou sólido) e a presença de partículas (tamanho e dispersão) devem ser consideradas, visto que a quantidade de água e distribuição das partículas podem afetar o tipo e a intensidade de absorção em um determinado comprimento de onda no espectro (ELY; THOMMES; CARVAJAL, 2008; PASQUINI, 2018).

A presença de água aumenta o número de ligações de hidrogênio, o que por sua vez diminui a intensidade de absorção da ligação química principal e desloca seu máximo para um comprimento de onda maior (PASQUINI, 2018). Segundo Barnes, Dhanoa e Lister (1989), a natureza física das partículas presentes nas amostras, bem como seu tamanho, influenciam, respectivamente, no espalhamento e na variação do caminho ótico do feixe, configurando duas das três principais causas de variação do espectro de refletância difusa no NIR.

No leite das diferentes espécies, enquanto suspensão coloidal e apresentando-se como líquido opaco e turvo, as principais partículas, sob o aspecto quantitativo, são as micelas de caseína e os glóbulos de gordura (HOLROYD, 2013; SINGH; GALLIER, 2017; CHEEMA et al., 2018). Estas partículas, por sua vez, causam grande espalhamento da luz e tornam o leite uma matriz desafiadora para a aplicação de NIRS (HOLROYD, 2013). O tamanho e a composição dessas partículas podem variar de acordo com a espécie, raça, estágio de lactação e com características do manejo (por exemplo, alimentação e quantidade de ordenhas diárias) (SINGH; GALLIER, 2017; CHEEMA et al., 2018; MA et al., 2019).

Cattaneo et al. (2009) demonstraram que o teor de gordura é o fator mais importante no espalhamento de radiação em amostras de leite fluido, enquanto o tamanho e a distribuição dos glóbulos contribuem para o espalhamento de forma menos significativa. Ainda nesse estudo, os autores relataram que o tamanho do glóbulo é mais importante que sua composição no que diz respeito ao espalhamento de radiação. Apesar disso, ainda são necessários estudos que estabeleçam um melhor entendimento sobre o impacto do tamanho e da distribuição dos glóbulos de gordura no espalhamento de radiação no espectro NIR (HOLROYD, 2013). Enquanto isso, a contribuição das

micelas de caseína no espalhamento de radiação em leite é baixa, devido a sua estrutura, propriedades parcialmente hidrofílicas, solubilidade e capacidade de suspensão (CATTANEO et al., 2009; AERNOUTS et al., 2015).

Além da variabilidade intrínseca da matéria-prima *in natura*, o processamento do produto também pode contribuir para a diferenciação das amostras. As etapas de processamento dependem do tipo do produto desejado e da planta de processamento (CHAVAN et al., 2016). Pequenos produtores e grandes indústrias podem obedecer a um mesmo protocolo de obtenção de um determinado produto ou acrescentar ou eliminar determinada etapa, a fim de adequar o processo a sua planta de processamento.

A homogeneização é uma das etapas frequentemente suprimida em pequenas usinas de beneficiamento de leite, enquanto é mais comum nas grandes indústrias. Tendo em vista que a homogeneização visa reduzir o tamanho dos glóbulos de gordura e prevenir sua separação durante o armazenamento (AERNOUTS et al., 2015; CHAVAN et al., 2016), esta etapa pode contribuir para produtos com características físicas diferenciadas.

Aernouts et al. (2015) demonstraram que a redução do tamanho dos glóbulos de gordura obtida por homogeneização resulta num menor espalhamento de radiação e num coeficiente de espalhamento mais dependente do comprimento de onda em amostras de leite fluido (maior em comprimentos de onda menor que 600 nm e menores em comprimentos de onda superiores).

Chen et al. (2019) observaram que, em leite de cabra homogeneizado, a distribuição do tamanho das micelas de caseína é mais uniforme e as micelas são menores e mais esféricas que as de leite não homogeneizado.

Ainda neste contexto, para leite em pó, a etapa da secagem (inerente ao processamento deste tipo de produto) pode gerar partículas com composições e tamanhos diferentes a depender de características do leite (perfis proteico e lipídico e quantidade de sólidos) (HABTEGEBRIEL et al., 2018; O'SULLIVAN et al., 2019) e das condições de operação do equipamento (vazão, tamanho do bico atomizador, temperaturas de entrada e saída) (O'SULLIVAN et al., 2019). Assim, para este tipo de produto, ainda são necessários estudos que verifiquem o impacto da distribuição do tamanho dos grânulos na variação do espectro NIR (HOLROYD, 2013).

Frequentemente, em estudos sobre a adulteração de leite de diferentes espécies e seus derivados, as amostras são provenientes de uma única fonte, tais como, produtores locais (DVORAK; MLCEK; SUSTOVA, 2016; MABOOD et al., 2017a; MABOOD et

al., 2017b), institutos de pesquisa (DVORAK; MLCEK; SUSTOVA, 2016) ou comércio local (produtos industrializados) (FU et al., 2014; HUANG et al., 2015; HUANG et al., 2016; LIM et al., 2016; FORCHETTI e POPPI, 2017), de forma que pouca variabilidade física é contemplada. Vale ressaltar que, quando a amostra é leite em pó, apenas alguns autores citam o tamanho das partículas, tanto da amostra quanto do adulterante, principalmente no caso de NIR hiperespectral (HUANG et al., 2016).

Já no que diz respeito às características operacionais de apresentação da amostra, a temperatura pode ser destacada como um interferente potencial significativo, pois sua alteração afeta diretamente a absorbância, causando deslocamento de bandas de absorção e variação na intensidade do sinal (POJIC; MASTILOVIC; MAJCEN, 2012; CAMPOS et al., 2018).

Shahul et al. (2018) verificaram que o aumento da temperatura causa variação no espectro NIR de amostras de leite fluido a 11, 25 e 59 °C. Neste estudo, os autores determinaram um desvio de 7% na absorção do comprimento de onda de 970 nm (banda de absorção de água), que foi atribuído ao enfraquecimento das ligações de hidrogênio. Além disso, verificou-se também um aumento do espalhamento com o aumento da temperatura, relacionado com alterações do índice de refração.

Segundo Beliciu e Moraru (2009), a alteração do índice de refração associada à temperatura pode ser correlacionada com a variação do tamanho das micelas de caseína sob diferentes temperaturas, considerando que os resultados do estudo demonstraram que há uma redução no tamanho das micelas com o aumento da temperatura (na faixa entre 6 e 50 °C).

A despeito deste fato, inúmeros trabalhos não controlam a temperatura das amostras, fato que pode influenciar a capacidade preditiva do modelo desenvolvido, principalmente em experimentos que visam determinar a umidade dos alimentos (POJIC; MASTILOVIC; MAJCEN, 2012; CAMPOS et al., 2018).

Nos estudos sobre adulteração de leite e derivados a temperatura das amostras não é comumente citada, principalmente se o objeto de estudo é o leite, quer seja líquido ou em pó (ZHANG et al., 2014; HUANG et al., 2016; LIM et al., 2016; MABOOD et al., 2017a; KARUNATHILAKA et al., 2018; LIU et al., 2018). Já quando a amostra é um derivado, apenas a temperatura de armazenamento é citada (BARRETO et al., 2018).

2.3.3 Variabilidade química das amostras

Ainda neste contexto, sabendo que para construir um modelo de calibração eficiente é necessário selecionar o grupo de amostras mais representativo possível, de forma que todas as fontes esperadas de variabilidade sejam consideradas (PASQUINI, 2018), a variabilidade química intrínseca das amostras biológicas, como é o caso do leite, pode ser destacada como a fonte mais relevante de variância em estudos sobre adulteração.

Para fins de sistematização, é possível agrupar as fontes de variabilidade da composição do leite (vide seção 2.2) que estão relacionadas às características do animal e ao manejo do rebanho em: quantitativas e qualitativas.

Podem ser classificados como quantitativos fatores como: período de lactação, quantidade de parições, condições ambientais (estação do ano) e idade dos animais, os quais afetam principalmente a concentração dos nutrientes presentes no leite (MASSOURAS; TRIANTAPHYLLOPOULOS; THEODOSSIOU, 2017; NAGY et al., 2017; NAGY et al., 2019), o que impacta na intensidade de absorção em um determinado comprimento de onda (COPPA et al., 2014; PASQUINI, 2018).

Já os fatores qualitativos, tais como, raça (perfil genético), localização geográfica (vegetação), alimentação e sistema de produção, estão associados a composições químicas diferenciadas, principalmente no que diz respeito aos perfis proteico e lipídico do leite produzido (VIEITEZ et al., 2016; MANOUSIDIS et al., 2018; KOTSAMPASI et al., 2017; SANT'ANA et al., 2019), o que pode influenciar na gama de comprimentos de onda que absorvem energia (COPPA et al., 2014; WANG; BOVENHUIS, 2018; PASQUINI, 2018).

Essa busca pela variabilidade amostral está bem evidente em estudos com objetivo de autenticação, determinação de origem e composição centesimal da matriz láctea. Nesses estudos, o número de amostras costuma ser elevado e proveniente de diferentes fontes e períodos do ano, buscando verificar a influência de diversos fatores, tais como alimentação (suplementação e biodiversidade de pasto), estágio de lactação e sazonalidade (WU; HE; FENG, 2008; ALTIERI et al., 2016; NUNEZ-SANCHEZ et al., 2016; REVILLA et al., 2017; ROZA-DELGADO et al., 2017; LIU et al., 2018; SUÁREZ et al., 2018).

Por outro lado, diferentes estudos que se propõem a determinar adulterantes em leite e derivados utilizam uma série de amostras que não contemplam tais fontes de variabilidade, de forma que não é possível determinar a influência do efeito matriz na precisão e na predição do modelo desenvolvido. Na maior parte dos estudos, amostras de leite e derivados provenientes de uma única fonte (raça ou período do ano) são adicionadas a diferentes concentrações da substância adulterante para construir o conjunto de amostras de calibração (FU et al., 2014; HUANG et al., 2015; DVORAK; MLCEK; SUSTOVA, 2016; HUANG et al., 2016; LIM et al., 2016; FORCHETTI e POPPI, 2017; MABOOD et al., 2017a; MABOOD et al., 2017b). Até o momento, apenas um trabalho tem contemplado a variabilidade amostral no momento da calibração do modelo (KARUNATHILAKA et al., 2018). Vale ressaltar ainda, que quando a substância adulterante é o leite de outra espécie, há ainda que se considerar igualmente as fontes de variabilidade do adulterante.

Outra fonte de variabilidade raramente considerada nos estudos sobre determinação de adulterantes em leite e derivados são as transformações químicas ocorridas durante o processamento da amostra. Mesmo em processamentos simples que incluem apenas tratamento térmico da amostra, como nos casos da pasteurização, esterilização e secagem, as alterações causadas, principalmente no que diz respeito às proteínas e aos glóbulos de gordura, devem ser avaliadas. Dentre as principais reações que ocorrem durante o tratamento térmico é possível citar a agregação e as mudanças conformacionais das proteínas, bem como reações de interação entre as proteínas e outras macromoléculas, como no caso da reação de escurecimento não enzimático. Tais reações, por sua vez, podem alterar o padrão de absorção no espectro NIR (PABARI et al., 2015; AALAEI; RAYNER; SJÖHOLM, 2016; MILORADOVIC et al., 2018; YE et al., 2019). Além disso, como citado anteriormente, outras etapas de produção, tais como homogeneização e armazenamento, podem influenciar no tamanho e na composição do glóbulo de gordura (HABTEGEBRIEL et al., 2018; MA et al., 2019; O’SULLIVAN et al., 2019; REN et al., 2019). Neste contexto, fica evidente a necessidade de verificar se as condições de processamento, bem como o momento da adulteração, antes ou depois do processamento, influenciam na capacidade preditiva do modelo desenvolvido.

2.3.4 Análise de dados (Quimiometria)

A extração das informações relevantes a partir dos dados gerados pela NIRS é o grande desafio ligado a este tipo de metodologia, visto que, na maioria das vezes, essas informações estão distribuídas por todo o espectro e não estão prontamente disponíveis para a determinação analítica (BRASCA; GOICOECHEA; CULZONI, 2018; PASQUINI et al., 2018).

Para tanto, é necessária a aplicação de ferramentas matemáticas e estatísticas aos dados de origem química, agrupadas sob o termo Quimiometria, que atendam a dois objetivos principais: extrair a informação analítica contida no espectro NIR, através da redução da dimensionalidade dos dados, sem grandes perdas de informações, extraindo assim as informações mais relevantes de uma série de dados, e prever propriedades relevantes em amostras desconhecidas (HUANG; ROMERO-TORRES; MOSHGBAR, 2010; BRASCA; GOICOECHEA; CULZONI, 2018; MANLEY; BAETEN, 2018; PASQUINI et al., 2018).

Nas análises químicas baseadas na NIRS, quer seja no modo refletância ou transmitância, as variações físicas da amostra, tais como tamanho e distribuição das partículas, densidade, formato e superfície da amostra são importantes fatores que afetam as medidas. Como consequência, as medidas realizadas em matrizes que possuem uma variabilidade nas características físicas podem produzir variações nos espectros sem correlação com as características químicas das amostras, prejudicando a precisão dos modelos desenvolvidos. Portanto, é imperativa a necessidade da aplicação de pré-processamentos para minimizar tais efeitos físicos e, consequentemente, tornar o efeito químico mais significativo (HUANG; ROMERO-TORRES; MOSHGBAR, 2010).

Diferentes ferramentas de pré-processamentos dos espectros são utilizadas para minimizar a influência das características físicas das amostras, dentre as quais destacam-se as derivadas de Savitzky-Golay, variação normal padrão (SNV, do inglês *standard normal variate*), correção do espalhamento multiplicativo (MSC, do inglês *multiplicative scatter correction*) e correção de linha de base (ELY; THOMMES; CARVAJAL, 2008; HUANG; ROMERO-TORRES; MOSHGBAR, 2010; KARUNATHILAKA et al., 2018). Os princípios, vantagens e desvantagens, bem como

suas indicações de utilização encontram-se bem descritas na literatura (SAVITZKY; GOLAY, 1964; HUANG; ROMERO-TORRES; MOSHGBAR, 2010).

As derivadas são as ferramentas mais comuns de pré-processamento de dados, sendo utilizadas para resolver sobreposição de pico, aumentar a resolução de pequenas diferenças entre os espectros das amostras e eliminar alterações constantes e lineares da linha de base causadas pelo espalhamento (SAVITZKY; GOLAY, 1964; HUANG; ROMERO-TORRES; MOSHGBAR, 2010)

A SNV é outro método frequentemente usado para pré-processamento dos dados devido à simplicidade do algoritmo e efetividade na correção do espalhamento. Este método é usado em espectros onde mudanças na linha de base e comprimento do caminho óptico causam diferenças entre os espectros (HUANG; ROMERO-TORRES; MOSHGBAR, 2010; MANLEY; BAETEN, 2018)

Outro método de pré-processamento que tem sido utilizado para minimizar os efeitos multiplicativos do espalhamento e as compensações de linha de base é o MSC. A normalização MSC é obtida por centralização na média (ou seja, cada espectro é subtraído da média da absorbância de cada comprimento de onda) e pela correção do espectro usando a inclinação do ajuste linear (HUANG; ROMERO-TORRES; MOSHGBAR, 2010; MANLEY; BAETEN, 2018).

Em estudos sobre adulteração de leite e derivados os autores frequentemente usam uma ou várias ferramentas de pré-processamento em combinação, tendo em vista as características complementares das mesmas, ou comparam a aplicação de diferentes métodos para alcançar os melhores índices de precisão do modelo (HUANG et al., 2015; LIM et al., 2016; WU et al., 2016; MABOOD et al., 2017a; MABOOD et al., 2017b; KARUNATHILAKA et al., 2018). Apenas em casos específicos, os autores dispensam o uso do pré-processamento por não observarem melhoras no modelo de classificação (WU et al., 2016; CHEN et al., 2017). É importante destacar que, nesses estudos há uma baixa variabilidade amostral (provenientes de uma única amostra comercial), de forma que a influência dos interferentes de natureza física é igualmente distribuída entre as amostras.

Essa hipótese pode ser mais evidentemente corroborada se forem considerados os diferentes estudos que analisam leite, com objetivo de classificar, determinar origem ou composição centesimal e, para tanto, utilizam um elevado número de amostras de origens diversas com o objetivo de contemplar a maior variabilidade possível. Nesses casos, a utilização de mais de uma ferramenta de pré-processamento de forma

combinada é imprescindível (ALTIERI et al., 2016; REVILLA et al., 2017; LIU et al., 2018).

Vale ressaltar, ainda, que apesar das ferramentas de pré-processamento serem eficientes na minimização de interferências de natureza física (e.g. tamanho dos glóbulos de gordura em leite fluido e tamanho das partículas em leite em pó), seu uso pode resultar em perdas de informações relevantes na diferenciação das amostras e, por este motivo, deve ser mais bem avaliado (WU; HE; FENG, 2008; CATTANEO et al., 2009; KARUNATHILAKA et al., 2018).

De forma análoga, na construção dos modelos preditivos, um ou mais algoritmos são aplicados para obter modelos mais robustos (HUANG et al., 2015; LIM et al., 2016; WU et al., 2016; MABOOD et al., 2017a; MABOOD et al., 2017b; KARUNATHILAKA et al., 2018).

Dentre os métodos mais utilizados para a extração da informação propriamente dita, a Análise de Componentes Principais (PCA, do inglês *Principal Component Analysis*) e a Regressão por Mínimos Quadrados Parciais (PLS, do inglês *Partial Least Squares*) se destacam enquanto métodos qualitativo e quantitativo, respectivamente.

A PCA é uma ferramenta para análise exploratória de dados multivariados, na qual uma série de variáveis correlacionadas é convertida em uma nova série de variáveis não correlacionadas (isto é, ortogonais) chamadas de componentes principais. A PCA redistribui a variância total de uma série de dados de forma que o primeiro componente principal explica a maior parte da variância, seguido do segundo componente e assim por diante (MABOOD et al., 2017a). Classificada como um método qualitativo ou de reconhecimento de padrão não supervisionado, a PCA é utilizada para reconhecer similaridades entre os dados espectrais e, posteriormente, identificar os agrupamentos naturais das amostras em diferentes classes (MANLEY; BAETEN, 2018).

Já o PLS é utilizado para relacionar uma ou mais variáveis resposta com diversas variáveis independentes, baseada no número de fatores (variáveis latentes), quer seja para fins de classificação ou de quantificação. Esses fatores são combinações lineares das variáveis resposta que melhor modelam as variáveis independentes, ou seja, estabelecem, tanto quanto possível, uma relação linear entre as projeções dos dados espectrais e as concentrações do(s) analito(s) (HUANG; ROMERO-TORRES; MOSHGBAR, 2010; MANLEY; BAETEN, 2018).

O número de fatores ou variáveis latentes deve ser escolhido cuidadosamente com a finalidade de capturar a maior quantidade possível de informação relacionada à

resposta. Nesse processo é necessário considerar que o uso de muitos fatores pode introduzir ruídos e informações não relevantes, gerando um modelo sobreajustado, enquanto o uso de poucos fatores pode desconsiderar informações relevantes resultando em um modelo pouco ajustado (CHEN et al., 2018).

Nesse contexto, com o objetivo de selecionar o conjunto de informações mais relevantes e obter modelos mais ajustados e com maior interpretabilidade, muitos autores avaliam a utilização de algoritmos de seleção de variáveis individuais ou por intervalos combinados ao PLS, para eliminação das informações não correlacionadas à resposta (WANG et al., 2018c; YUN et al., 2019). Algoritmo genético (GA, do inglês *genetic algorithm*), seleção de variáveis passo atrás (BVS, do inglês *backward variable selection*), eliminação de variável não-informativa (UVE, do inglês *uninformative variable elimination*), seleção de intervalo único (iPLS, do inglês *interval selection PLS*) e o algoritmo das projeções sucessivas (SPA, do inglês *successive projection algorithm*) são alguns dos algoritmos usados com esse objetivo. Eles têm demonstrado ser vantajosos em termos de ajustes dos modelos construídos a partir de dados espectrais (HE et al., 2016; WU et al., 2016; DINIZ; PISTONESI; ARAÚJO, 2015; CHEN et al., 2018; KREPPER et al., 2018).

Sob este aspecto, também é possível verificar que estudos que não consideram as principais fontes de variabilidade amostral, utilizando um pequeno número de amostras, observam uma distribuição mais estreita ao longo das componentes principais, quando comparadas a estudos que utilizam amostras provenientes de fontes diversas, que costumam evidenciar distribuições mais amplas (WU et al., 2016; CHEN et al., 2017; MABOOD et al., 2017a; MABOOD et al., 2017b; KARUNATHILAKA et al., 2018). Além disso, de forma geral, em estudos com baixa variabilidade amostral, os autores não precisam recorrer à utilização de ferramentas de seleção de variáveis para obter modelos ajustados (MABOOD et al., 2017a; MABOOD et al., 2017b).

Por outro lado, para fins quantitativos, é preciso considerar que, sem a representatividade adequada das fontes de variância, a aplicabilidade e a capacidade preditiva dos modelos desenvolvidos são comprometidas.

2.4 IMAGENS DIGITAIS

Uma imagem é uma representação gráfica de um objeto tridimensional em uma superfície plana, ou seja, uma representação bidimensional. Adicionalmente, a imagem digital é o arranjo bidimensional de dados com valores de intensidade e cor para cada elemento ou pixel, que, por sua vez, representam cada coordenada espacial da imagem. Este arranjo bidimensional constitui uma matriz de números. O tamanho desta matriz é determinado pela resolução espacial da imagem e, portanto, quanto maior a matriz menor será o tamanho de cada pixel (DANIEL, 2009).

A cor e a intensidade de luz de cada pixel são representadas digitalmente por 8 bits assumindo valores discretos entre 0 e 255 (ou seja, correspondendo a valores de 2^8), totalizando 256 tons/variáveis. Além disso, a informação sobre a cor também pode variar de acordo com a quantidade de canais existentes. Na imagem em preto e branco, por exemplo, há apenas um canal, no qual a cor é expressa em vários tons de cinza, enquanto que alguns sistemas, tais como o RGB (vermelho, verde e azul, do inglês, *red, green, blue*), possuem 3 canais e expressam a cor através da combinação destes, resultando em 768 variáveis possíveis (DANIEL, 2009; MCANDREW, 2015).

Além desses, outros espaços baseados em intensidade-cromaticidade têm sido utilizados para representar uma imagem, tais como HSI (matiz, saturação, intensidade), HSB (matiz, saturação, brilho) e HSL (matiz, saturação, luminosidade). A cromaticidade é definida como o grau de pureza da cor e está relacionada com matiz e saturação, em que matiz representa o tipo de cor (por exemplo, azul, magenta, verde, amarelo e etc.) e saturação se refere ao quanto a cor está diluída com branco. (MCANDREW, 2015).

Para extração das informações através das imagens digitais são utilizados histogramas de cor, que por sua vez descrevem a distribuição da frequência dos pixels em função de cada componente de cor. Considerando que a cor é uma propriedade aparente atribuída a distribuição de um espectro de luz, os histogramas dependem da fonte de iluminação, a qual emite a luz que interage com o objeto (amostra), que por sua vez é registrada pelo dispositivo de captura de imagem (DINIZ et al., 2012; MCANDREW, 2015; FERNANDES et al., 2019).

Para construir um histograma de cor é necessário, inicialmente, calcular quantos valores pertencem a cada uma das variáveis possíveis. Essas variáveis são geralmente

especificadas como consecutivas e consequentemente de tamanhos iguais, em vez de intervalos de variáveis em histogramas tradicionais. O eixo vertical é então a frequência de distribuição, isto é, o número de ocorrências de cada variável no eixo horizontal (MCANDREW, 2015).

Neste contexto, os diferentes tons de cor registrados nas imagens digitais podem ser relacionados com a composição complexa da matriz amostral, de tal forma que o histograma de cor pode ser explorado como uma impressão digital deste tipo de amostra. Como o histograma de cada componente de cor (escala de cinza, R, G, B, H, S, I, V, B, S e L) é composto por 256 tons, a natureza multivariada da informação analítica pode ser acessada pelo uso combinado ou individual desses histogramas.

A principal vantagem do uso dos histogramas de cor combinado com ferramentas quimiométricas é a aquisição rápida de informação analítica sem a necessidade de processamento da imagem adicional além da extração do histograma (FERNANDES et al., 2019).

Diferentes dispositivos têm sido utilizados para obtenção de imagens digitais com fins analíticos, tais como, câmeras digitais de modelos comerciais (DINIZ et al., 2012; EL-SAYED et al., 2016; FERNANDES et al., 2019), câmeras de smartphones (BOTELHO; DANTAS; SENA, 2017; AZEEM; MADBOULY; EL-SHARAT, 2019), scanners (BOTELHO; ASSIS; SENA, 2014) e webcams (COSTA et al., 2015). Consequentemente, as metodologias baseadas em imagens digitais exibem elevado potencial de acessibilidade aos analistas.

As informações obtidas através das imagens digitais têm sido amplamente empregadas em diversas áreas da ciência, tais como, medicina, veterinária, agronomia (DANIEL, 2009; AL-JANABI et al., 2012; ALGANCI et al., 2014). Na ciência e tecnologia dos alimentos (CTA), a utilização das imagens digitais está comumente associada ao atributo cor e, por este motivo, é frequentemente utilizada no acompanhamento do amadurecimento de frutos, por exemplo (MANNINEN et al., 2015; EL-SAYED et al., 2016).

Por outro lado, estudos recentes na área de CTA têm ampliado o uso desta técnica para outras matérias-primas e para outras finalidades, tais como, aspectos de qualidade em carnes (LARRAÍN; SCHAEFER; REED, 2008), análise de frações da composição centesimal (ROMANO et al., 2012; CRUZ-FERNÁNDEZ et al., 2017; FERNANDES et al., 2019), detecção do uso de aditivos (BOTELHO; ASSIS; SENA, 2014; BOTELHO; DANTAS; SENA, 2017; AZEEM; MADBOULY; EL-SHARAT,

2019) e rastreabilidade por origem geográfica (DINIZ et al., 2012; PIERINI et al., 2016).

No que diz respeito à matriz leite, o emprego de imagens digitais também tem alcançado destaque, tendo sido utilizada tanto para determinação de constituintes naturais do leite (KUCHERYAVSKIY; MELENTEVA; BOGOMOLOV, 2014) quanto para detectar a presença de contaminantes (SANTOS, WENTZELL e PEREIRA-FILHO, 2012; MASAWAT; HARFIELD; NAMWONG, 2015).

Tendo em vista que o uso isolado de imagens digitais como metodologia analítica está relacionado com o atributo físico de cor, é importante considerar os fatores que influenciam este atributo em leites caprino e bovino. Em contexto amplo, a cor branca do leite deve-se a reflexão da luz pelas micelas de caseína em todos os comprimentos de onda, de tal forma que, quanto maior a concentração de caseínas, mais branco será o leite, visto que estas reduzem a penetração da luz. Os glóbulos de gordura também são responsáveis pela dispersão da luz, entretanto, contribuem de forma menos significativa para a cor branca do leite (MISAWA; BARBANO; DRAKE, 2016).

Em bovinos, a cor do leite pode variar de acordo com aspectos genéticos e de manejo, tais como raça, estágio de lactação, horário de ordenha, estação do ano, sistema de alimentação e condições sanitárias (SERT et al., 2016; MCDERMOTT et al., 2016; SCARSO et al., 2017). Neste contexto, diferentes autores estabeleceram uma correlação positiva entre teor de gordura, concentração de β -caroteno e a cor amarelada em leite das diferentes raças bovinas (PHILLIPS et al., 1995; NOZIÈRE et al., 2006; MCCARTHY; LOPETCHARAT; DRAKE, 2017). Por outro lado, leites de origem caprina não apresentam β -caroteno em sua composição, resultado da total conversão dessa substância em vitamina A, o que se reflete nos teores mais elevados dessa vitamina em leite caprino quando comparado ao leite bovino (RAYNAL-LJUTOVAC et al., 2008; CLAEYS et al., 2014). Assim, tanto o leite caprino quanto seus derivados tendem a apresentar uma coloração mais branca que os de origem bovina (WANG; ZHU; WANG, 2016).

Outra fonte de variabilidade da cor dos leites das diferentes espécies é o processamento. De forma geral, a cor do leite varia de acordo com a temperatura de processamento, tornando-se mais escuro à medida que se eleva a temperatura. Tal fato está correlacionado à intensificação das reações de escurecimento não enzimático, principalmente quando os valores ultrapassam os 100 °C, como no caso da esterilização e da secagem (REDDY et al., 2014; DEVI et al., 2015).

3 MATERIAL E MÉTODOS

3.1 AMOSTRAS

As amostras de leite caprino foram obtidas mediante parceria com pequenos produtores das Regiões de Campina Grande e do Cariri Paraibano, como detalhado no **Quadro 2**. As amostras foram obtidas a partir de diferentes produtores, no período entre novembro de 2017 e outubro de 2019.

Todos os rebanhos eram constituídos de animais de diferentes idades e períodos de lactação. Não houve controle da alimentação, a fim de garantir a variabilidade química real da matéria-prima. O sistema de manejo variou entre extensivo, semi-extensivo e intensivo.

As amostras foram coletadas, mediante acompanhamento da ordenha, e transportadas sob refrigeração em caixa isotérmica, para o Núcleo de Pesquisa e Extensão em Alimentos (NUPEA) da Universidade Estadual da Paraíba (UEPB) em Campina Grande. Imediatamente após a chegada ao laboratório, as amostras foram pasteurizadas a 65 °C por 30 minutos e resfriadas a 4 °C.

De forma análoga, as amostras de leite fluido bovino foram obtidas de pequenos produtores das regiões de Campina Grande, Boa Vista, Boqueirão, Sumé, Soledade, Serra Branca, Taperoá e Sobrado na Paraíba e Cachoeirinha em Pernambuco. Os procedimentos de obtenção, transporte, processamento e armazenamento foram os mesmos empregados ao leite caprino.

Quadro 2- Descrição das amostras utilizadas na identificação e quantificação de adulteração de leite caprino por adição de leite bovino, por espectroscopia NIR e imagens digitais.

Ano de coleta	Lote	Tipo de criação	Raça	Período de coleta	Localização geográfica	Armazenamento
2017	Bovino 1	Semi-extensivo	Girolando	Novembro	Boa vista/PB	Congelamento
	Bovino 2	Semi-extensivo	Girolando	Dezembro	Cachoeirinha/PE	Congelamento
2018	Caprino 1	Semi-extensivo	Mestiças	Janeiro	Cabaceiras/ PB	Congelamento
	Caprino 2	Semi-extensivo	Mestiças	Março	Cabaceiras/PB	Congelamento
	Caprino 3	Semi-extensivo	Mestiças	Abril	Cabaceiras/PB	Congelamento
	Caprino 4	Extensivo	Mestiças	Junho	Cabaceiras/PB	Resfriamento
	Caprino 5	Extensivo	Mestiças	Junho	Cabaceiras/PB	Resfriamento
	Caprino 6	Extensivo	Mestiças	Junho	Cabaceiras/PB	Resfriamento
	Caprino 7	Extensivo	Mestiças	Junho	Cabaceiras/PB	Resfriamento
	Bovino 3	Extensivo	Girolando	Junho	Boqueirão/PB	Resfriamento
	Bovino 4	Extensivo	Mestiça	Junho	Soledade/PB	Resfriamento
	Bovino 5	Extensivo	Mestiça	Junho	Sumé/PB	Resfriamento
	Bovino 6	Extensivo	Mestiça	Junho	Serra Branca/PB	Resfriamento
	Bovino 7	Extensivo	Girolando	Junho	Campina Grande/PB	Resfriamento
2019	Caprino 8	Semi-extensivo	Moxotó	Outubro	Taperoá/PB	Resfriamento
	Caprino 9	Semi-extensivo	Moxotó	Outubro	Taperoá/PB	Resfriamento
	Caprino 10	Semi-extensivo	Murciana	Outubro	Taperoá/PB	Resfriamento
	Caprino 11	Semi-extensivo	Alpina Americana	Outubro	Taperoá/PB	Resfriamento
	Caprino 12	Semi-extensivo	Parda	Outubro	Taperoá/PB	Resfriamento
	Caprino 13	Intensivo	Anglo Nubiana	Outubro	Campina Grande /PB	Resfriamento
	Bovino 8	Intensivo	Sindi	Outubro	Taperoá/PB	Resfriamento
	Bovino 9	Intensivo	Guzerá	Outubro	Taperoá/PB	Resfriamento
	Bovino 10	Semi-extensivo	Sindi/ Guzerá	Outubro	Taperoá/PB	Resfriamento
	Bovino 11	Extensivo	Girolando	Outubro	Sobrado/PB	Resfriamento

3.2 EXPERIMENTOS

3.2.1 Experimento 1- Equipamento NIR de bancada

O primeiro experimento foi realizado utilizando 7 lotes de leite de caprino, 7 lotes de leite bovino (Caprino 1 a 7 e Bovino 1 a 7 – **Quadro 2**), bem como suas misturas, analisadas em equipamento NIR de bancada (item 3.4). Para detalhamento das características e das possíveis fontes de variabilidade das amostras do experimento 1, tais amostras estão descritas a seguir:

Leite caprino – Dos 7 lotes analisados, 4 eram provenientes de um mesmo produtor do município de Cabaceiras, coletados nos meses de janeiro, março, abril e junho de 2018, e os demais lotes foram coletados de diferentes produtores, também da região de Cabaceiras (raio de 20 km), em um mesmo período (junho de 2018). Todos os rebanhos eram constituídos de cabras mestiças. A coleta dos leites caprinos foi realizada nas fazendas de origem.

Leite bovino – Todos os 7 lotes analisados foram obtidos de diferentes produtores situados em diferentes localidades das regiões de Campina Grande, Boa Vista, Boqueirão, Sumé, Soledade e Serra Branca, na Paraíba e Cachoeirinha em Pernambuco, sendo 4 lotes de vacas puras da raça Girolando e 3 lotes de vacas mestiças. 2 lotes foram coletados em 2017 (novembro e dezembro) e os demais (5) em junho de 2018. A coleta dos leites bovinos foi realizada em Campina Grande, uma vez que os animais das diferentes localidades foram transportados para feira municipal de comercialização de animais.

3.2.2 Experimentos 2 (Equipamento NIR portátil) e 3 (Imagens Digitais)

O segundo e o terceiro experimento foram realizados utilizando 6 lotes de leite de caprino, 4 lotes de leite bovino (Caprino 8 a 13 e Bovino 8 a 11 – **Quadro 2**), bem como suas misturas, analisadas em equipamento NIR portátil (item 3.4) e em scanner (item 3.5), respectivamente. Para detalhamento das características e das possíveis fontes de variabilidade das amostras dos experimentos 2 e 3, tais amostras estão descritas a seguir:

Leite caprino – Dos 6 lotes analisados, 5 eram provenientes de um mesmo município (Taperoá-PB), e 1 proveniente do município de Campina Grande. Todos os lotes foram coletados no mesmo período (outubro de 2019). Os rebanhos eram constituídos de cabras puras das raças: Alpina Americana, Anglo Nubiana, Moxotó, Murciana e Parda. A coleta dos leites caprinos foi realizada nas fazendas de origem.

Leite bovino – Dos 4 lotes analisados, 3 foram obtidos de um mesmo produtor, porém de diferentes raças (Sindi e Guzerá), submetidas a diferentes sistemas de criação (intensivo e semi-extensivo), situado no município de Taperoá-PB. O lote remanescente foi proveniente de um produtor localizado no município de Sobrado-PB, de animais da

raça Girolando, em sistema de criação semi-extensivo. Todos os lotes foram adquiridos em um mesmo período (outubro de 2019). A coleta dos leites bovinos foi realizada nas fazendas de origem.

3.3 OBTENÇÃO DAS AMOSTRAS ADULTERADAS DE LEITE CAPRINO

As amostras adulteradas de leite caprino foram obtidas mediante mistura individual e aleatória dos diferentes lotes de leites caprino e bovino puros em diferentes níveis. Para o experimento 1, no qual as amostras foram analisadas em equipamento NIR de bancada, os níveis avaliados foram: 1, 3, 5, 7, 10, 20, 25,30, 40, 50 e 75% de leite bovino (m/m) em um total de 20 g cada. Já nos experimentos 2 e 3, nos quais as amostras foram analisadas no equipamento NIR portátil e em scanner, os níveis avaliados foram: 5, 10, 15, 20, 30, 40, 50, 60 e 75 % de leite bovino (m/m) em um total de 10 g cada.

A homogeneização foi realizada em tubos tipo Falcon de 50 mL por sucessivas inversões. As análises foram realizadas em duplicatas. As amostras foram dispostas em placas de petri lisas de poliestireno 60 x 15 mm, para posterior aquisição de espectros NIR e imagens digitais.

3.4 OBTENÇÃO DOS ESPECTROS NIR

Nos experimentos 1 e 2, espectros NIR foram adquiridos em triplicata, à temperatura ambiente (23 ± 1 °C), em dois equipamentos diferentes: i) **Experimento 1**- Equipamento de bancada FT-NIR (4000 – 10000 cm^{-1}), modelo Analyzer Diamond 20, marca Applied Instrument Technologies no modo reflectância difusa, e ii) **Experimento 2** – Equipamento NIR portátil (900 a 1650 nm), modelo DLP NIRscan Nano Evaluation Module, marca Texas Instruments. Cada espectro foi obtido a partir de uma média de 32

varreduras. O espectro de uma pastilha de politetrafluoretileno (PTFE) foi usado como branco.

3.5 AQUISIÇÃO DAS IMAGENS DIGITAIS

As imagens digitais das amostras foram obtidas utilizando o scanner de uma impressora multifuncional HP, modelo Deskjet 4180 com 300 dpi de resolução, gerando uma matriz de dimensões 2550 x 3507 pixels. A imagem de cada amostra foi adquirida individualmente, em uma mesma área do scanner, para padronizar os efeitos da incidência da luz. As imagens foram obtidas em triplicata, a partir das quais foram obtidos histogramas médios de cada amostra a partir de uma região de interesse (ROI, do inglês, *region of interest*) circular localizada no centro de cada imagem. Os histogramas nos sistemas de cor escala de cinza (GS, do inglês, *grayscale*), vermelho, verde e azul (RGB) e matiz, saturação e intensidade (HSI, do inglês, *hue, saturation, intensity*) foram obtidos a partir de cada imagem digital usando a interface gráfica Image_GUI, escrita em ambiente Matlab e disponibilizada gratuitamente em: http://laqa.quimica.ufpb.br/Softwares%20para%20Download/Imagens_gui.rar.

Considerando que cada componente de cor é representado digitalmente por 8 bits, os quais permitem que estes assumam valores entre 0 e 255, totalizando 256 variáveis, um vetor (x) de informação instrumental, com dimensões de 1 x 1792, contemplando todos os componentes de cor avaliados neste trabalho, foi obtido para cada amostra. Os histogramas de cor foram avaliados individualmente e em combinação.

3.6 ANÁLISE QUIMIOMÉTRICA DOS DADOS

Para o tratamento dos dados obtidos por NIRS dos experimentos 1 e 2, foram utilizadas ferramentas matemáticas e estatísticas que empregam métodos quimiométricos de pré-processamento, classificação e calibração multivariada, dependendo da finalidade de aplicação. As diferentes ferramentas foram testadas e

aplicadas para redução do ruído e seleção de regiões espectrais mais informativas, objetivando a melhoria da capacidade preditiva dos modelos.

Como os espectros apresentam espalhamento e variações sistemáticas da linha de base, diferentes técnicas de pré-processamento foram aplicadas aos dados espectrais: SNV, MSC, filtro média móvel, correção linear de linha de base e/ou *offset*. Diferentes combinações das técnicas supracitadas também foram avaliadas. A suavização por média móvel foi realizada com 7, 9, 11 e 13 pontos.

Todos os procedimentos quimiométricos foram implementados em Matlab® 2019a (Mathworks, USA).

3.6.1 Tratamento dos dados - Experimento 1

No que diz respeito aos experimentos conduzidos no equipamento NIR de bancada, das 148 amostras analisadas, 18 eram de leite caprino puro (7 lotes originais e suas replicatas), 18 eram de leite bovino puro (7 lotes originais e suas replicatas), e as demais (112) eram amostras adulteradas (misturas de leite caprino e bovino em diferentes níveis).

Neste conjunto de dados, para fins de classificação, PCA e PLS-DA foram utilizados como técnicas de reconhecimento de padrão não supervisionada e supervisionada, respectivamente. Para o modelo PLS-DA, cada classe foi dividida em conjunto de treinamento (70%) e teste (30%), aplicando algoritmo Kennard-Stone (KS), que leva em consideração as distâncias de x e y simultaneamente, de forma que as amostras com menores e maiores valores de y são obrigatoriamente incluídas na série de treinamento, evitando, desta forma, problemas de extrapolação. O modelo PLS-DA foi construído e então validado utilizando validação cruzada completa, enquanto as amostras do conjunto de teste foram utilizadas somente para avaliação final da performance do modelo, em termos de precisão, sensibilidade e especificidade.

A precisão foi calculada como o número de classificações corretas dividido pelo total de amostras do conjunto (treinamento ou teste). Já a sensibilidade foi calculada como o número de amostras corretamente classificadas na classe dividido pelo número total de amostras pertencentes àquela classe, enquanto especificidade foi calculada como o número de decisões negativas corretas (negativos verdadeiros) dividido pelo

total de amostras negativas (negativos verdadeiros e falso positivos) (MASSART et al., 1998; LAVINE, 2009).

Para fins de quantificação, amostras com níveis acima de 1% de leite bovino (amostras adulteradas e amostras de leite bovino puro) foram consideradas para a construção do modelo de calibração. Assim, 90 e 40 amostras foram selecionadas para os conjuntos de calibração e predição, respectivamente, aplicando algoritmos KS. Então, três diferentes algoritmos PLS foram utilizados para a construção de modelos de calibração: (a) PLS convencional utilizando toda a informação espectral (*full-spectrum* PLS) e suas variações acopladas com técnicas de seleção de intervalos: *i*PLS (*Interval* PLS), onde apenas um intervalo é utilizado, e *i*SPA-PLS (PLS combinado com algoritmo de sucessivas projeções para seleção de intervalos), quando diferentes combinações de intervalos podem ser selecionadas.

O número de fatores PLS (variáveis latentes, VL) e o número de intervalos foram otimizados de acordo com a raiz do erro quadrático médio da validação cruzada (RMSECV, do inglês, *root mean square error of crossvalidation*) e o coeficiente de correlação da validação cruzada (r_{cv}) no conjunto de calibração. A predição foi somente utilizada para a avaliação final dos dados e comparação dos modelos de calibração. A habilidade preditiva do modelo final foi avaliada em termos dos valores de RMSEP (raiz do erro quadrático médio da predição, do inglês, *root mean square error of prediction*), r_{pred} (coeficiente de correlação da predição) e relação de desempenho do desvio (RPD_{pred} , do inglês, *ratio performance deviation*) no conjunto de predição externa, o qual não foi empregado no processo de construção do modelo.

3.6.2 Tratamento dos dados – Experimento 2

Para os experimentos conduzidos no equipamento NIR portátil, cujo objetivo principal foi identificar a presença de adulteração (classificação) independentemente do montante de leite bovino adicionado, apenas duas classes de amostras foram consideradas: a) amostras de leite caprino puro, constituídas dos 6 lotes originais e de combinações aleatórias entre eles, resultando em 54 amostras e b) amostras adulteradas, constituídas de misturas de leite caprino e bovino em diferentes níveis, resultando em 146 amostras.

Para tanto, os algoritmos de classificação OC-PLS (do inglês, *One-class – Partial least squares*), PLS para análise discriminante (PLS-DA, do inglês, *PLS-Discriminant analysis*) e *i*SPA-PLS-DA foram empregados e então comparados em termos de precisão, sensibilidade e especificidade.

Para o modelo OC-PLS, 40 amostras de leite caprino puro foram selecionadas por KS como conjunto de treinamento. As demais 14 amostras de leite caprino puro foram incluídas no grupo de teste com as 146 amostras adulteradas, para avaliar a robustez do modelo.

Para os modelos PLS-DA e *i*SPA-PLS-DA, o algoritmo KS foi utilizado para selecionar 70 e 30% das amostras de cada classe (puras e adulteradas) para os conjuntos de treinamento e teste, respectivamente. Desta forma, os conjuntos de treinamento e teste foram constituídos, respectivamente, de 38 e 16 amostras para a classe de amostras puras e de 102 e 44 amostras para a classe de amostras adulteradas.

3.6.3 Tratamento dos dados – Experimento 3

Assim como para o Experimento 2, no Experimento 3, o objetivo principal foi identificar a presença da adulteração de leite caprino por adição de leite bovino, independentemente do montante de adulterante presente na amostra, e por este motivo apenas duas classes de amostras foram consideradas: a) amostras de leite caprino puro, constituídas dos 6 lotes originais e de combinações aleatórias entre eles, resultando em 54 amostras e b) amostras adulteradas, constituídas de misturas de leite caprino e bovino em diferentes níveis, resultando em 138 amostras.

Para avaliar a capacidade de identificação de amostras puras e adulteradas de leite caprino através dos dados adquiridos por imagens digitais, foram aplicados os algoritmos de classificação “one-class” OCPLS e DD-SIMCA (do inglês, *Data driven-Soft independent modelling of class analogies*) cujas performances foram avaliadas em termos de eficiência, sensibilidade e especificidade. A eficiência foi calculada como a média geométrica da sensibilidade e da especificidade (OLIVERI; DOWNEY, 2012).

Para os modelos construídos pelos algoritmos OC-PLS e DD-SIMCA, que consideram apenas amostras da classe alvo, 40 amostras de leite caprino puro foram selecionadas por KS como conjunto de treinamento. As 14 amostras de leite caprino

puro restantes foram incluídas no grupo de teste, com as 138 amostras adulteradas, para avaliar a robustez do modelo.

3.7 ANÁLISES FÍSICO-QUÍMICAS DAS AMOSTRAS DE LEITES CAPRINO E BOVINO

As análises físico-químicas dos leites foram realizadas no Núcleo de Pesquisa e Extensão em Alimentos (NUPEA) da UEPB - Campus I em Campina Grande/PB.

As determinações do teor de lipídeos das amostras puras de leite foram realizadas utilizando o método Gerber (2000.18), de acordo com o recomendado pelos métodos oficiais da AOAC. Já o teor de proteínas foi estimado pela determinação do conteúdo total de nitrogênio obtido através do método micro-Kjedhal e multiplicando o resultado pelo fator de conversão (6.38) de acordo com método oficial (690.52 e 991.20) (AOAC, 2005).

REFERÊNCIAS

- AALAEI, K.; RAYNER, M.; SJÖHOLM, I. Storage stability of freeze-dried, spray-dried and drum-dried skim milk powders evaluated by available lysine. **LWT- Food Science and Technology**, v. 73, p. 675-682, 2016.
- AERNOOTS, B.; VAN BEERS, R.; WATTÉ, R.; HUYBRECHTS, T.; JORDENS, J.; VERMEULLEN, D.; VAN GERVEN, T.; LAMMERTYN, J.; SAEYS, W. Effect of ultrasonic homogenization on the VIS/NIR bulk optical properties of milk. **Colloids and Surfaces B: Biointerfaces**, v. 126, p. 510-519, 2015.
- AGRIMONTI, C.; PIRONDINI, A.; MARMIROLI, M.; MARMIROLI, N. A quadruplex PCR (qPCR) assay for adulteration in dairy products. **Food Chemistry**, v. 187, p. 58-64, 2015.
- AHMED, A. S.; EL-BASSIONY, T.; ELMALT, L. M.; IBRAHIM, H. R. Identification of potent antioxidant bioactive peptides from goat milk proteins. **Food Research International**, v. 74, p. 80-88, 2015.
- ALBENZIO, M.; CAMPANOZZI, A.; D'APOLITO, M.; SANTILLO, A.; MANTOVANI, M. P.; SEVI, A. Differences in protein fraction from goat and cow milk and their role on cytokine production in children with cow's milk protein allergy. **Small Ruminant Research**, v. 105, p. 202-205, 2012.
- ALGANCI, U.; OZDOGAN, M.; SERTEL, E.; ORMECI, C. Estimating maize and cotton yield in southeastern Turkey with integrated use of satellite images, meteorological data and digital photographs. **Field Crops Research**, v. 157, p. 8-19, 2014.
- ALICHANIDIS, E.; MOATSOU, G.; POLYCHRONIADOU, A. Composition and properties of non-cow milk and products. *In: Non-bovine and milk products*, Elsevier, p. 81-116, 2016.
- AL-JANABI, S.; HUISMAN, A.; WILLEMS, S. M.; VAN DIEST, P. J. Digital slide images for primary diagnostics in breast pathology: a feasibility study. **Human Pathology**, v. 43, p. 2318-2325, 2012.
- ALMEIDA, O. C.; FERRAZ JR, M. V. C.; SUSIN, I.; GENTIL, R. S.; POLIZEL, D. M.; FERREIRA, E. M.; BARROSO, J. P. R.; PIRES, A. V. Plasma and milk fatty acid

profiles in goats fed diets supplemented with oils from soybean, linseed or fish. **Small Ruminant Research**, v. 170, p. 125-130, 2019.

ALTIERI, G.; GENOVESE, F.; ADMANE, N.; DI RENZO, G. C. On-line measure of donkey's milk properties by near infrared spectrometry. **LWT- Food Science and Technology**, v. 69, p. 348-357, 2016.

AOAC (2005). Official Methods of Analysis of Association of Official Analytical Chemists. 18th Edition, Washington, DC.

AZEEM, S. M. A.; MADBOULY, M. D.; EL-SHARAT, M. F. Determination of nitrite in processed meat using digital image method and powdered reagent. **Food Composition Analysis**, v. 81, p. 28-36, 2019.

BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. **Applied Spectroscopy**, v. 43, n. 5, 1989.

BARRETO, A.; CRUZ-TIRADO, J. P.; SICHE, R.; QUEVEDO, R. Determination of starch content in adulterated fresh cheese using hyperspectral imaging. **Food Bioscience**, v. 21, p. 14-19, 2018.

BELICIU, C. M.; MORARU, C. I. Effect of solvent and temperature on the size distribution of casein micelles measured by dynamic light scattering. **Journal of Dairy Science**, v. 92, n. 5, p. 1829-1839, 2009.

BILGE, G.; SEZER B.; ESELLER, K. E.; BERBEROGLU, H.; TOPCU, A.; BOYACI, I. H. Determination of whey adulteration in milk powder by using laser induced breakdown spectroscopy. **Food Chemistry**, v. 212, p. 183-188, 2016.

BOTELHO, B. G.; ASSIS, L. P.; SENA, M. M. Development and analytical validation of a simple multivariate calibration method using digital scanner images for sunset yellow determination in soft beverages. **Food Chemistry**, v. 159, p. 175-180, 2014.

BOTELHO, B. G.; DANTAS, K. C. F.; SENA, M. M. Determination of allura red dye in hard candies by using digital images obtained with a mobile phone and N-PLS. **Chemometrics and Intelligent Laboratory Systems**, v. 167, p. 44-49, 2017.

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

BRASCA, R.; GOICOECHEA, H. C.; CULZONI, M. J. Multiway calibration approaches for quality control of food sample. *In: Food Safety and preservation: Modern biological approaches to improving consumer health*, Academic Press, p. 143-165, 2018.

CAMPOS, M.I.; ANTOLIN, G.; DEBÁN, L.; PARDO, R. Assessing the influence of temperature on NIRS prediction models for the determination of sodium content in dry-cured ham slices. **Food Chemistry**, v. 257, p. 237-242, 2018.

CAROPRESE, M.; CILIBERTI, M. G.; SANTILLO, A.; MARINO, R.; ALBENZIO, M. Immune response, productivity and quality of Milk from grazing goats as affected by dietary polyunsaturated fatty acid supplementation. **Research in Veterinary Science**, v. 105, p. 229-235, 2016.

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

CATTANEO, T. M. P.; CABASSI, G. PROFAIZER, M. GIANGIACOMO, R. Contribution of light scattering to near infrared absorption in milk. **Journal of Near Infrared spectroscopy**, v. 17, p. 337-343, 2009.

CEBALLOS, L. S.; MORALES, E. R.; ADARVE, G. T.; CASTRO, J. D.; MARTÍNEZ, L. P.; SAMPELAYO, M. R. S. Composition of goat and cow milk produced under similar conditions and analyzed by identical methodology. **Journal of Food Composition and Analysis**, v. 22, p. 322-329, 2009.

CEBO, C.; LOPEZ, C.; HENRY, C.; BEAUVALLET, C.; MÉNARD, O.; BEVILACQUA, C.; BOUVIER, F.; CAILLAT, H.; MARTIN, P. Goat α s1-casein genotype affects milk fat globule physicochemical properties and the composition of the milk fat globule membrane. **Journal of Dairy Science**, v. 95, p. 62215-62229, 2012.

CHAVAN, R. S.; SEHRAWAT, R.; MISHRA, V.; BHATT, S. B. Milk: Processing of milk. **Encyclopedia of Food and Health**, 2016.

CHEEMA, M.; SMITH, P. B.; PATTERSON, A. D.; HRISTOV, A.; MARTE, F. M. The association of lipophilic phospholipids with native bovine casein micelles in skim

milk: Effect of lactation stage and casein micelle size. **Journal of Dairy Science**, v. 101, p. 1-16, 2018.

CHEN, R. K.; CHANG, L. W.; CHUNG, Y. Y.; LEE, M. H.; LING, Y. C. Quantification of cow adulteration in goat milk using high-performance liquid chromatography with electrospray ionization mass spectrometry. **Rapid Communications in mass spectrometry**, v. 18, p. 1167-1171, 2004.

CHEN, B., LEWIS, M. J., GRANDISON, A. S. Effect of seasonal variation on the composition and properties of raw milk destined for processing in the UK. **Food Chemistry**, v. 158, p. 216-223, 2014.

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, 2017.

CHEN, H.; TAN, C.; LIN, Z.; WU, T. Classification and quantification of milk powder by near-infrared spectroscopy and mutual information-based variable selection and partial least squares. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 189, p. 183-189, 2018.

CHEN, D.; LI, X. Y.; ZHAO, X.; QIN, Y. S.; ZHANG, X. X.; LI, J.; WANG, J. M.; WANG, C. F. Proteomics and microstructure profiling of goat milk protein after homogenization. **Journal of Dairy Science**, v. 102, p. 1-12, 2019.

CHILLIARD, Y.; TORAL, P. G.; SHINGFIELD, K. J.; ROUEL, J.; LEROUX, C.; BERNARD, L. Effects of diet and physiological factors on milk fat synthesis, milk fat composition and lipolysis in the goat: A short review. **Small Ruminant Research**, v. 122, p. 31-37, 2014.

CHOOPAN, R.; THANAKIATKRAI, P.; KITPIPIT, T. Simultaneous species identification in milk and dairy products using direct pcr. **Forensic Science International: Genetix Supplement Series**, v.6, p. 214-215, 2017.

CLAEYS, W. L.; VERRAES, C.; CARDOEN, S.; DE BLOCK, J.; HUYGHEBAERT, A.; RAES, K.; DEWETTINCK, K.; HERMAN, L. Consumption of raw or heated milk from different species: An evaluation of the nutritional and potential health benefits. **Food Control**, v.42, p. 188-201, 2014.

CLARK, S.; GARCÍA, M. B. M. A 100-year review: Advances in goat milk research. **Journal of Dairy Science**, v. 100, p. 10026-10044, 2017.

COPPA, M.; REVELLO-CHION, A.; GIACCONE, D.; FERLAY, A.; TABACCO, E.; BORREANI, G. Comparison of near and medium infrared spectroscopy to predict fatty acid composition on fresh and thawed milk. **Food Chemistry**, v. 150, p. 49-57, 2014.

COSTA, W. K. A.; SOUZA, E. L. D.; BELTRÃO-FILHO, E. M.; VASCONCELOS, G. K. V.; SANTI-GADELHA, T.; GADELHA, C. A. A.; FRANCO, O. L.; QUEIROGA, R. C. R. E.; MAGNANI, M. Comparative Protein Composition Analysis of Goat Milk Produced by the Alpine and Saanen Breeds in Northeastern Brazil and Related Antibacterial Activities. **PLoS ONE**, v. 9, n. 3, e93361, 2014.

COSTA, G. B.; FERNANDES, D. D. S.; ALMEIDA, V. E.; ARAÚJO, T. S. P.; MELO, J. P.; DINIZ, P. H. G. D.; VÉRAS, G. Digital image-based classification of biodiesel. **Talanta**, v. 139, p. 50-55, 2015.

CRUZ-FERNÁNDEZ, M.; LUQUE-COBIJA, M. J.; CERVERA, M. L.; MORALES-RUBIO, A.; LA GUARDIA, M. Smartphone determination of fat in cures meat products. **Microchemical Journal**, v. 132, p. 8-14, 2017.

DABROWSKA, A.; WALECKA, E.; BANIA, J. ZELAZKO, M.; SZOLTSIC, M.; CHRZANOWSKA, J. Quality of UHT goat's milk in Poland evaluated by real-time PCR. **Small Ruminant Research**, v. 94, p. 32-37, 2010.

DANIEL, G. B. Digital Imaging. **Veterinary Clinics of North America: Small Animal Practice**, v. 39, p. 667-676, 2009.

DAVE, A. C.; SINGH, H. Milk protein interaction. *In: Encyclopedia of Food Chemistry*, Elsevier, p. 63-69, 2019.

DEETH, H.; BANSAL, N. Whey proteins: an overview. *In: Whey Proteins*, Elsevier, p. 1-50, 2019.

DEVI, A. F.; BUCKOW, R.; SING, T.; HEMAR, Y.; KASAPIS, S. Colour change and proteolysis of skim milk during high pressure thermal-processing. **Journal of Food Engineering**, v. 147, p. 102-110, 2015.

DIAS, L. A.; PERES, A. M.; VELOSO, A. C. A.; REIS, F. S.; VILAS-BOAS, M.; MACHADO, A. A. S. C. Na electronic tongue taste evaluation: Identification of goat milk adulteration with bovine milk. **Sensors and Actuators B: Chemical**, v. 136, p. 209-217, 2009.

DINIZ, P. H. G. D.; DANTAS, H. V.; MELO, K. D. T.; BARBOSA, M. F.; HARDING, D. P.; NASCIMENTO, E. C. L.; PISTONESI, M. F.; BAND, B. S. F.; ARAÚJO, M. C. U. Using a simple digital camera and SPA-LDA modeling to screen teas. **Analytical Methods**, v. 4, p. 2648-2652, 2012.

DINIZ, P.H.G.D.; PISTONESI, M.F.; ARAÚJO, M.C.U. Using iSPA-PLS and NIR spectroscopy for the determination of total polyphenols and moisture in commercial tea samples. **Analytical Methods**, v. 7, p. 3379–3384 , 2015.

DI PINTO, A.; TERIO, V.; MARCHETTI, P.; BOTTARO, M.; MOTTOLA, A.; BOZZO, G.; BONERBA, E.; CECI, E.; TANTILLO, G. DNA-based approach for species identification of goat-milk products. **Food Chemistry**, v. 229, p. 93-97, 2017.

DUPONT, D.; CROGUENNEC, T.; POCHET, S. Milk Proteins – Analytical Methods. *In. Reference Module in Food Science*, Elsevier, 2018.

DVORAK, L.; MLCEK, J.; SUSTOVA, K. Comparison of FT-NIR spectroscopy and ELISA for detection of adulteration of goat cheeses with cow's Milk. **Food Composition and Additives**, v. 99, n. 1, p. 180-186, 2016.

EUROPEAN COMMISSION 2018/150. Regulation (EU) 2016/1240 as regards methods for the analysis and quality evaluation of milk and milk products eligible for public intervention and aid for private storage. **Official Journal of the European Commission**, L26/14, 2018.

EGITO, A. S.; ROSINHA, G. M. S.; LAGUNA, L. E.; MICLO, L.; GIRARDET, J. M.; GAILLARD, J. L. Fast eletrophoretic detection method of adulteration of caprine milk by bovine milk. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 58, n. 5, p. 932-939, 2006.

EL-HATMI, H.; JRAD, Z.; SALHI, I.; AGUIBI, A.; NADRI, A.; KHORCHANI, T. Comparison of composition and whey protein fractions of human, camel, donkey, goat and cow milk. **Mljekarstvo**, v. 65, n. 3, p. 159-167, 2015.

EL-SAYED, S.; GALAL, H.; ALLAN, A.; SCHMIDHALTER, U. Passive reflectance sensing and digital image analysis for assessing quality parameters of mango fruits. **Scientia Horticulturae**, v. 212, p. 136-147, 2016.

EL-TARABANY, M. S.; EL-TARABANY, A. A.; ROUSHDY, E. M. Impact o lactation stage on milk composition and blood biochemical and hematological parameters of dairy Baladi goats. **Saudi Journal of Biological Sciences**, 2016.

ELY, D. R.; THOMMES, M.; CARVAJAL, M.T. Analysis of the effects of particle size and densification on NIR spectra. **Colloids and Surfaces A: Physicochemical and Engineering Aspects**, v. 331, p. 63-67, 2008.

FAOSTAT. Food and Agriculture Organization of the United Nations Statistics Division. 2018. Disponível em: <http://faostat3.fao.org/browse/Q/QL/E>. Acesso em: 12 de janeiro de 2019.

FELISBERTO, N. R. O.; OLIVEIRA, L. S.; CORDEIRO, A. G. P. C. Sistemas de produção de caprinos leiteiros. *In*: Workshop sobre produção de caprinos na região da mata atlântica, 13. Coronel Pacheco. **Anais... Sobral: Embrapa Caprinos e Ovinos**; Coronel Pacheco: Embrapa Gado de Leite, p. 11-35, 2016.

FERNANDES, D.D.S.; ROMEO, F.; KREPPER, G.; DI NEZIO, M.S.; PISTONESI, M.F.; CENTURIÓN, M.E.; ARAÚJO, M.C.U.; DINIZ, P.H.G.D. Quantification and identification of adulteration in the fat content of chicken hamburgers using digital images and chemometric tools. **LWT - Food Science and Technology**, v. 100, p. 20–27, 2019.

FINETE, V. L. M.; GOUVÊA, M. M.; MARQUES, F. F. C.; PEREIRA NETTO, A. D. Is it possible to screen for milk or whey protein adulteration with melamine, urea and ammonium sulphate, combining Kjeldahl and classical spectrophotometric methods? **Food Chemistry**, v. 141, p. 3649-3655, 2013.

FORCHETTI, D. A. P.; POPPI, R. J. Use of NIR hyperspectral imaging and multivariate curve resolution (MCR) for detection and quantification of adulterants in milk powder. **LWT- Food Science and Tecnology**, v. 76, p. 337-343, 2017.

FU, X.; KIM, M. S.; CHAO, K.; QIN, J.; LIM, J.; LEE, H.; GARRIDO-VARO, A.; PÉREZ-MARÍN, D.; YING, Y. Detection of melamine in milk powders based on NIR hyperspectral imaging and spectral similarity analyses. **Journal of Food Engineering**, v. 127, p. 97-104, 2014.

GENIS, D. O.; SEZER, B.; BILGE, G.; DURNA, S.; BOYACI, I. H. Development of synchronous fluorescence method for identification of cow, goat, ewe and buffalo milk species. **Food Control**, v. 108, 1006808, 2020.

GINGER, M. R.; GRIGOR, M. R. Comparative aspects of milk caseins. **Comparative Biochemistry and Physiology Part B**, v. 124, p. 133-145, 1999.

GOLINELLI, L. P.; CARVALHO, A. C.; CASAES, R. S.; LOPES, C. S. C.; DELIZA, R.; PASCHOALIN, V. M. F.; SILVA, J. T. Sensory analysis and species-specific PCR detect bovine milk adulteration of frescal (fresh) goat cheese. **Journal of Dairy Science**, v. 97, p. 6693-6699, 2014.

GÓMEZ-RUIZ, J. A.; MIRALLES, B.; AGUERA, P.; AMIGO, L. Quantitative determination of α s2- and α s1-casein in goat's milk with different genotypes by capillary electrophoresis. **Journal of Chromatography A**, v. 1054, p. 279-284, 2004.

GONZÁLEZ-MARTÍN, M. I.; PALACIOS, V. V.; REVILLA, I.; VIVAR-QUINTANA, A. M.; HERNÁNDEZ-HIERRO, J. M. Discrimination between cheeses made from cow's, ewe's, goat's milk from unsaturated fatty acids and use of the canonical biplot method. **Journal of Food Composition and Analysis**, v. 56, p. 34-40, 2017.

HABTEGEBRIEL, H.; EDWARD, D.; WAWIRE, M.; SILA, D.; SEIFU, E. Effect of operating parameters on the surface and physic-chemical properties of spray-dried camel milk powders. **Food and Bioproducts Processing**, v. 112, p. 137-149, 2018.

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

HE, S.; FANG, S.; LIU, X.; ZHANG, W.; XIE, W.; ZHANG, H.; WEI, D.; FU, W.; PEI, D. Investigation of a genetic algorithm based cubic spline smoothing for baseline correction of Raman spectra. **Chemometrics and Intelligent Laboratory Systems**, v. 152, p. 1-9, 2016.

HOLROYD, S. E. The use of near infrared spectroscopy on milk and milk products. **Journal of Near Infrared Spectroscopy**, v. 21, p. 311-322, 2013.

HOLT, C. Casein and casein micelle structures, functions and diversity in 20 species. **International Dairy Journal**, v. 60, p. 2-13, 2016.

HUANG, M.; KIM, M. S.; DELWICHE, S. R.; CHAO, K.; QIN, J.; MO, C.; ESQUERRE, C.; ZHU, Q. Quantitative analysis of melamine in milk powders using near-infrared hyperspectral imaging and band ratio. **Journal of Food Engineering**, v. 181, p. 10-19, 2016.

HUANG, Y.; TIAN, K.; MIN, S.; XIONG, Y.; DU, G. Distribution assessment and quantification of counterfeit melamine in powdered milk by NIR imaging methods. **Food Chemistry**, v. 177, p. 174-181, 2015.

HUANG, J.; ROMERO-TORRES, S.; MOSHGBAR, M. Practical considerations in data pré-processing for NIR and Raman spectroscopy. **American Pharmaceutical Review**, 2010.

IBGE. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Censo Agropecuário 2017 – Resultados preliminares. **Censo agropecuário**, v.7, p. 1-108, 2018.

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

KAROUI, R.; DE BAERDEMAEKER, J. 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, p. 621-640, 2007.

KARUNATHILAKA, S. R.; YAKES, B. J.; HE, K.; CHUNG, K. K.; MASSOBA, M. Non-targeted NIR spectroscopy and SIMCA classification for commercial milk powder authentication: A study using eleven potential adulterants. **Heliyon**, v. 4, e00806, 2018.

KHOLIF, A. E.; MORSY, T. A.; ABDO, M. M. Crushed flaxseed versus flaxseed oil in the diets of Nubian goats: Effect on feed intake, digestion, ruminal fermentation, blood chemistry, milk production, milk composition and milk fatty acid profile. **Animal Feed Science and Technology**, v. 244, p. 66-75, 2018.

KIM, M. J.; KIM, H. Y. Direct duplex real-time loop mediated isothermal amplification assay for the simultaneous of cow and goat species origin of milk and yogurt products for field use. **Food Chemistry**, v. 246, p. 26-31, 2018.

KOTSAMPASI, B.; CHRISTODOULOU, C.; TSIPLAKOU, E.; MAVROMMATIS, A.; MITSIOPOULOU, C.; KARAIKOU, C.; DOTAS, V.; ROBINSON, P. H.; BAMPIDIS, V. A.; CHRISTODOULOU, V.; ZERVAS, G. Effects of dietary pomegranate pulp silage supplementation on milk yield and composition, milk fatty acid profile and blood plasma antioxidant status of lactating dairy cows. **Animal Feed science and Technology**, v. 234, p. 228-236, 2017.

KREPPER, G.; ROMEO, F.; FERNANDES, D.D.S.; DINIZ, P.H.G.D.; ARAÚJO, M.C.U.; DI NEZIO, M.S.; PISTONESI, M.F.; CENTURIÓN, M.E. Determination of fat content in chicken hamburgers using NIR spectroscopy and the Successive Projections Algorithm for interval selection in PLS regression (iSPA-PLS). **Spectrochimica Acta A**, v. 189, p. 300–306, 2018.

KUCHERYAVSKIY, S.; MELENTEVA, A.; BOGOMOLOV, A. Determination of fat and total protein content in milk using conventional digital imaging. **Talanta**, v. 121, p. 144-152, 2014.

LARRAÍN, R. E.; SCHAEFER, D. M.; REED, J. D. Use of digital images to estimate CIE color coordinates of beef. **Food Research International**, v. 41, p. 380-385, 2008.

LAVINE, B. K. Validation of classifiers. In Walczak B, Tauler R, Brown S (Vol.Eds.), *Comprehensive chemometrics*: Vol. 3, (pp. 587–599). Oxford: Elsevier. 2009.

LI, Q.; ZHAO, Y.; ZHU, D.; PANG, X.; LIU, Y.; FREW, R.; CHEN, G. Lipidomics profiling of goat milk, soymilk and bovine milk by UPLC-Q-Exactive orbitrap mass spectrometry. **Food Chemistry**, v. 224, p. 309-309, 2017.

LIM, J.; KIM, G.; MO, C.; KIM, M. S.; CHAO, K.; QIN, J.; FU, X.; BAEK, I.; CHO, B. K. Detection of melamine in milk powders using near-infrared hyperspectral imaging combined with regression coefficient of partial least square regression model. **Talanta**, v. 151, p. 183-191, 2016.

LIU, N.; PARRA, H. A.; PUSTJENS, A.; HETTINGA, K.; MONGONDRY, P.; RUTH, S. M. Evaluation of portable near-infrared spectroscopy for organic milk authentication. **Talanta**, v. 184, p. 128-135, 2018.

LOHUMI, S.; LEE, S.; LEE, H.; CHO, B. K. A review of vibrational spectroscopy techniques for the detection of food authenticity and adulteration. **Trends in Food Science & Technology**, v. 46, p. 85-98, 2015.

LÓPEZ-CALLEJA, I.; GONZÁLEZ, I.; FAJARDO, V.; RODRÍGUEZ, M. A.; HERNÁNDEZ, P. E.; GARCÍA, T.; MARTÍN, R. Rapid detection of cows' milk in sheeps' and goats' milk by a species-specific polymerase chain reaction technique. **Journal of Dairy Science**, v. 87, p. 2839-2845, 2004.

MA, Y.; ZHANG, L.; WU, Y.; ZHOU, P. Changes in milk fat globule membrane proteome after pasteurization in human, bovine and caprine species. **Food Chemistry**, v. 279, p. 209-215, 2019.

MABOOD, F.; JABEEN, F.; AHMED, M.; HUSSAIN, J.; AL MASHAYKHI, S. A. A.A.; AL RUBAIEY, Z. M. A; FAROOQ, S.; BOQUÉ, R.; ALI, L.; HUSSAIN, Z.; AL-HARASSI, 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, 2017a.

MABOOD, F.; JABEEN, F.; HUSSAIN, J.; AL-HARASSI, A.; HAMAED, A.; AL MASHAYKHI, S. A. A. A.; AL RUBAIEY, Z. M. A; MANZOOR, S.; KHAN, A.; HAQ, I. Q. M.; GILANI, S. A.; KHAN, A. FT-NIRS coupled with chemometric methods as a rapid alternative tool for the detection & quantification of cow milk adulteration in camel milk samples. **Vibrational Spectroscopy**, v. 92, p. 245-250, 2017b.

MANLEY, M; BAETEN, V. Spectroscopic Technique: Near Infrared (NIR) Spectroscopy. *In*. **Modern Techniques for Food Authentication**, 2ed, Academic Press, p. 51-102, 2018.

MANNINEN, H.; PAAKKI, M.; HOPIA, A.; FRANZÉN, R. Measuring the green color of vegetables from digital images using image analysis. **LWT- Food Science and Technology**, v.63, p. 1184-1190, 2015.

MANOUSIDIS, T.; PARISSI, Z. M.; KYRIAZOPOULOS, A. P.; MALESIOS, C.; KOUTROUBAS, S. D.; ABAS, Z. Relationships among nutritive value of selected forages, diet composition and Milk quality in goats grazing in a Mediterranean Woody rangeland. **Livestock Science**, v. 218, p. 8-19, 2018.

MAPA – MINISTÉRIO DA AGRICULTURA PECUÁRIA E ABASTECIMENTO. **Instrução Normativa Nº 37, de 31 de outubro de 2000.** Aprovar o Regulamento Técnico de Identidade e Qualidade de Leite de Cabra, conforme consta do Anexo desta Instrução Normativa.

MAPA – MINISTÉRIO DA AGRICULTURA PECUÁRIA E ABASTECIMENTO. **Instrução Normativa Nº 51, de 18 de setembro de 2002.** Aprova os Regulamentos Técnicos de Produção, Identidade e Qualidade do Leite tipo A, do Leite tipo B, do Leite tipo C, do Leite Pasteurizado e do Leite Cru Refrigerado e o Regulamento Técnico da Coleta de Leite Cru Refrigerado e seu Transporte a Granel, em conformidade com os Anexos a esta Instrução Normativa.

MAPA – MINISTÉRIO DA AGRICULTURA PECUÁRIA E ABASTECIMENTO. **Instrução Normativa Nº 68, de 12 de dezembro de 2006.** Oficializar os Métodos Analíticos oficiais físico-químicos, para controle de leite e produtos lácteos, em

conformidade com o Anexo desta Instrução Normativa, determinando que sejam utilizados nos laboratórios nacionais agropecuários.

MAPA – MINISTÉRIO DA AGRICULTURA PECUÁRIA E ABASTECIMENTO.
Instrução Normativa N° 76, de 26 de novembro de 2018a. Aprovar os Regulamentos Técnicos que fixam a identidade e as características de qualidade que devem apresentar o leite cru refrigerado, o leite pasteurizado, o leite pasteurizado tipo A, na forma desta Instrução Normativa e do Anexo único.

MAPA – MINISTÉRIO DA AGRICULTURA PECUÁRIA E ABASTECIMENTO.
Instrução Normativa N° 77, de 26 de novembro de 2018b. Estabelecer os critérios e procedimentos para produção, acondicionamento, conservação, transporte, seleção e recepção de leite cru em estabelecimentos registrados no serviço de inspeção oficial, na forma desta Instrução Normativa e do seu Anexo.

MASAWAT, P.; HARFIELD, A.; NAMWONG, A. An iPhone-based digital image colorimeter for detecting tetracycline in milk. **Food Chemistry**, v. 144, p. 23-29, 2015.

MASSART, D. L.; VANDEGINSTE, B. G. M.; BUYDENS, S. J., LEWI, P. J.; SMEYERS-VERBEKE, J. Handbook of chemometrics and qualimetrics. Amsterdam: Part A. Elsevier. 1998.

MASSOURAS, T.; TRIANTAPHYLLOPOULOS, K. A.; THEODOSSIOU, I. Chemical composition, protein fraction and fatty acid profile of donkey Milk during lactation. **International Dairy Journal**, v. 75, p. 83-90, 2017.

MCANDREW, A. A computational introduction to digital image processing. 2ª Ed, CRC Press, 2015.

MCCARTHY, K. S.; LOPETCHARAT, K.; DRAKE, M. A. Milk fat threshold determination and effect of milk fat content on consumer preference for fluid milk. **Journal of Dairy Science**, v. 100, p. 1702-1711, 2017.

MCDERMOTT, A.; VISENTIM, G.; MCPARLAND, E.; BERRY, D. P.; FENELON, M. A.; DE MARCHI, M. Effectiveness of mid-infrared spectroscopy to predict the color of bovine milk and the relationship between milk color and traditional milk quality traits. **Journal of Dairy Science**, v. 99, p. 3267-3273, 2016.

MCGHEE, C. E.; JONES, J. O.; PARK, Y. W. Evaluation of textural and sensory characteristics of three types of low fat goat milk ice cream. **Small Ruminant Research**, v. 123, p. 293-300, 2015.

- MCLEOD, R. J.; PROSSER, C. G.; WAKEFIELD, J. W. Identification of goat milk powder by manufacturer using multiple chemical parameters. **Journal of Dairy Science**, v. 99, n. 2, p. 982-993, 2016.
- MILORADOVIC, Z.; MIOCINOVIC, J.; KLJAJEVIC, N.; TOMASEVIC, I.; PUDJA, P. The influence of Milk heat treatment on composition, texture, colour and sensory characteristics of cows' and goats' Quark-type cheeses. **Small Ruminant Research**, Article in press, 2018.
- MISAWA, N.; BARBANO, D. M.; DRAKE, M. Influence of casein as a percentage of true protein and protein level on color and texture of milk containing 1 and 2% fat. **Journal of Dairy Science**, v. 99, p.1-21, 2016.
- MOLINA, E.; MARTÍN-ÁLVAREZ, P. J.; RAMOS, M. Analysis of cows', ewes' and goats' milk mixtures by capillary electrophoresis: quantification by multivariate regression analysis. **International Dairy Journal**, v. 9, p. 99-105, 1999.
- MONTALBANO, M.; SEGRETO, R.; DI GERLANO, R.; MASTRANGELO, S.; SARDINA, M. T. Quantitative determination of casein genetic variants in goat milk: Application in Girgentana dairy goat breed. **Food Chemistry**, v. 192, p. 760-764, 2016.
- MOORE, J. C.; SPINK, J.; LIPP, M. Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. **Journal of Food Science**, v. 77, n. 4, p. 118-126, 2012.
- NAGY, P.; FÁBRI, Z. N.; VARGA, L.; REICZIGEL, J.; JUHÁSZ, J. Effect of genetic and nongenetic factors on chemical composition of individual Milk samples from dromedary camels (*Camelus dromedarius*) under intensive management. **Journal of Dairy Science**, v. 100, p. 8680-8693, 2017.
- NAGY, P.; JUHÁSZ, J.; REICZIGEL, J.; CSÁSZÁR, G.; KOCSIS, R.; VARGA, L. Circannual changes in major chemical composition of bulk dromedary camel milk as determined by FT-MIR spectroscopy, and factors of variation. **Food Chemistry**, v. 278, p. 248-253, 2019.
- NASCIMENTO, C. F.; SANTOS, P. M.; PEREIRA-FILHO, E. R.; ROCHA, F. R. P. Recent advances on determination of milk adulterants. **Food Chemistry**, v. 221, p. 1232-1244, 2017.
- .

NIRWAL, S.; PANT, R.; RAI, N. Analysis of milk quality, adulteration and mastitis in milk samples collected from different regions of Dehradun. **International Journal of Pharm Tech Research**, v. 5, n. 2, p. 359-364, 2103.

NOZIÈRE, P.; GROLIER, P.; DURAND, D.; FERLAY, A.; PRADEL, P.; MARTIN, B. Variations in carotenoids, fat-soluble micronutrients, and color in cow's plasma and milk following changes in forage and feeding level. **Journal of Dairy Science**, v. 89, p. 2634-2648, 2006.

NÚÑEZ-SANCHEZ, N.; MARTÍNEZ-MARÍN, A. L.; POLVILLO, O.; FERNÁNDEZ-CABANÁS, V. M.; CARRIZOSA, J.; URRUTIA, B.; SERADILLA, J. L. Near Infrared Spectroscopy (NIRS) for the determination of the milk fat fatty acid profile of goats. **Food Chemistry**, v. 190, p. 244-252, 2016.

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

O'SULLIVAN, J. J.; NORWOOD, E. A.; O'MAHONY, J. A.; KELLY, A. L. Atomisation technologies used in spray drying in the dairy industry: A review. **Journal of Food Engineering**, v. 243, p. 57-69, 2019.

PABARI, R. M.; TOGASHI, D.; CAMA-MONCUNILL, R.; EL ARNAOUT, T.; RIFAI, H.; CRUISE, P.; CULLEN, J. P.; SULLIVAN, C. Multipoint near-infrared spectrometry for real-time monitoring of protein conformational stability in powdered infant formula. **Food Composition and Analysis**, v. 66, n. 5, p. 526-532, 2015.

PARK, Y. W. Hypo-allergenic and therapeutic significance of goat milk. **Small Ruminant Research**, v. 14, p. 151-159, 1994.

PASQUINI, C. Near Infrared Spectroscopy: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, v. 1026, p. 8-36, 2018.

PASQUINI, C. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. **Journal Brazilian Chemistry Society**, v. 14, n. 2, p. 198-212, 2003.

PERINELLI, D. R. BONACUCINA, G.; CESPI, M.; BONAZZA, F.; PALMIERI, G. F.; PUCCIARELLI, S.; POLZONETTI, V.; ATTARIAN, L.; POLIDORI, P.; VINCENZETTI, S. A comparison among β -caseins purified from milk of different

species: Self assembling behavior and immunogenicity potential. **Colloids and Surfaces B: Biointerfaces**, v. 173, p. 210-216, 2019.

PESIC, M.; BARAC, M.; VRVIC, M.; RISTIC, N.; MACEJ, O.; STANOJEVIC, S. Qualitative and quantitative analysis of bovine milk adulteration in caprine and ovine milks using native-PAGE. **Food Chemistry**, v. 125, n. 4, p. 1443-1449, 2011.

PHILLIPS, L. G.; MCGIFF, M. L.; BARBANO, D. M.; LAWLESS, H. T. The influence of fat on the sensory properties, viscosity, and color of lowfat milk. **Journal of Dairy Science**, v. 78, p. 1258, 1266, 1995.

PIERINI, G. D.; FERNANDES, D. D. S.; DINIZ, P. H. G. D.; ARAÚJO, M. C. U.; DI NEZIO, M. S.; CENTURIÓN, M. E. A digital image-based traceability tool of the geographical origins of argentine propolis. **Microchemical Journal**, v. 128, p. 62-67, 2016.

POJIC, M.; MASTILOVIC, J.; MAJCEN, N. Robustness of the near infrared spectroscopy method determined using univariate and multivariate approach. **Food Chemistry**, v. 134, p. 1699-1705, 2012.

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, p. 211-230, 2015.

RAFIQ, S.; HUMA, N.; PASHA, I.; SAMEEN, A.; MUKHTAR, O.; KHAN, M. I. Chemical composition, nitrogen fractions and amino acids profile of milk from different animal species. **Asian-Australasian Journal of Animal Sciences**, v. 29, n. 7, p. 1022-1028, 2016.

RAYNAL-LJUTOVAC, K.; LAGRIFFOUL, G.; PACCARD, P.; GUILLET, I.; CHILLIARD, Y. Composition of goat and sheep milk products: an update. **Small Ruminant Research**, v. 79, p. 57-72, 2008.

REDDY, R. S.; RAMACHANDRA, C. T.; HIREGOUDAR, S.; NIDONI, U.; RAM, J.; KAMMAR, M. Influence of processing conditions on functional and reconstitution properties of milk powder made from Osmanabadi goat milk by spray drying. **Small Ruminant Research**, v. 119, p. 130-137, 2014.

REN, Q. R.; ZHANG, H.; GUO, H. Y.; JIANG, L.; TIAN, M.; REN, F. Z. Detection of cow milk adulteration in yak milk by ELISA. **Journal of Dairy Science**, v. 97, p. 6000-6006, 2014.

REN, Q.; LI, L.; DUDU, O. E.; MA, Y. Thermal and structural changes of pasteurized milk fat globules during storage. **Food Bioscience**, v. 28, p. 27-35, 2019.

REVILLA, I.; ESCUREDO, O.; GONZÁLEZ-MARTÍN, M. I.; PALACIOS, C. Fatty acids and fat-soluble vitamins in ewe's milk predicted by near infrared reflectance spectroscopy. Determination of seasonality. **Food Chemistry**, v. 214, p. 468-477, 2017.

RODRIGUES, N. P. A.; GIVISIEZ, P. E. N.; QUEIROGA, R. C. R. E.; AZEVEDO, P. S.; GEBREYES, W. A.; OLIVEIRA, C. J. B. Milk adulteration: Detection of bovine milk in bulk goat milk produced by smallholders in northeastern Brazil by a duplex PCR assay. **Journal of Dairy Science**, v. 95, n. 5, p. 2749-2752, 2012.

RODRÍGUEZ, N.; ORTIZ, M. C.; SARABIA, L.; GREDILLA, E. Analysis of protein chromatography profiles joint to partial least squares to detect adulterations in milk mixtures and cheeses. **Talanta**, v. 81, p. 255-264, 2010.

ROMANO, G.; ARGYROPOULOS, D.; NAGLE, M.; KHAN, M. T.; MULLER, J. Combination of digital images and laser light to predict moisture content and color of bell pepper simultaneously during drying. **Journal of Food Engineering**, v. 109, p. 438-448, 2012.

ROZA-DELGADO, B.; GARRIDO-VARO, A.; SOLDADO, A.; ARROJO, A. G.; VALDÉS, M. C.; MAROTO, F.; PÉREZ-MARÍN, D. Matching portable NIRS instruments for *in situ* monitoring indicators of milk composition. **Food Control**, v. 76, p. 74-81, 2017.

SANT'ANA, A. M. S.; BESSA, R. J. B.; ALVES, S. P.; MEDEIROS, A. N.; COSTA, R. G.; BATISTA, A. S. M.; MADRUGA, M. S.; QUEIROGA, R. C. R. E. Fatty acid, volatile and sensory profiles of milk and cheese from goats raised on native semiarid pasture or in confinement. **International Dairy Journal**, v. 91, p. 147-154, 2019.

SANTOS, P. M.; WENTZELL, P. D.; PEREIRA-FILHO, E. R. Scanner digital images combined with color parameters: A case study to detect adulterations in liquid cow's milk. **Food Analytical Methods**, v. 5, p. 89-95, 2012.

SANTOS, P. M.; PEREIRA-FILHO, E. R. Digital images analysis –an alternative tool for monitoring milk authenticity. **Analytical Methods**, v. 5, p. 3669-3674, 2013.

SANTOS, J. R.; VIEGAS, O.; PÁSCOA, R. N. M. J.; FERREIRA, I. M. P. L. V. O.; RANGEL, A. O. S. S.; LOPES, J. A. In-line monitoring of the coffee roasting process

with near infrared spectroscopy: Measurement of sucrose and colour. **Food Chemistry**, v. 208, p. 103-110, 2016.

SATIR, G.; GUZEL-SEYDIM, Z. B. How kefir fermentation can affect product composition? **Small Ruminant Research**, v. 134, p. 1-7, 2016.

SAVITSZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627- 1639, 1964.

SCARSO, S.; MCPARLAND, S.; VISENTIM, G.; BERRY, D. P.; MCDERMOTT, A.; DE MARCHI, M. Genetic and nongenetic factors associated with milk color in dairy cows. **Journal of Dairy Science**, v. 100, p. 7345-7361, 2017.

SERHAN, M.; MATTAR, J.; DEBS, L. Concentrated yogurt (Labneh) made of mixture of goat's and cow's milk: Physicochemical, microbiological and sensory analysis. **Small Ruminant Research**, v. 138, p. 46-52, 2016.

SERT, D.; MERCAN, E.; AYDEMIR, S.; CIVELEK, M. Effects of milk somatic cell counts on some physicochemical and functional characteristics of skim and whole milk powders. **Journal of Dairy Science**, v. 99, p. 5254-5264, 2016.

SHAUL, O.; FANRAZI-KAHANA, M.; MEITAV, O.; PINHASI, G. A.; ABOOKASIS, D. Study the effects of varying interference upon the optical properties of turbid samples using NIR spatial light modulation. **Optics Communications**, v. 411, p. 70-79, 2018.

SHODER, D. Melamine milk powder and infant formula sold in east Africa. **Journal of Food Protection**, v. 73, n. 9, p. 1709-1714, 2010.

SILVA, P. D. L.; BEZERRA, M. F.; SANTOS, K. M.O.; CORREIA, R. T. P. Potentially probiotic ice cream from goat's Milk: Characterization e cell viability during processing, storage and simulated gastrointestinal conditions. **LWT- Food Science and Technology**, v. 62, p. 452-457, 2015.

SINGH, H.; GALLIER, S. Nature's complex emulsion: The fat globules of milk. **Food Hydrocolloids**, v. 68, p. 81-89, 2017.

SONG, H.; XUE, H.; HAN, Y. Detection of cow's milk in Shaanxi goat's milk with an ELISA assay. **Food Control**, v. 22, n. 6, p. 883-887, 2011.

SOUZA, S. S.; CRUZ, A. G.; WALTER, E. H. M.; FARIA, J. A. F.; CELEGHINI, R. M. S.; FERREIRA, M. M. C.; GRANATO, D.; SANT'ANA, A. S. Monitoring the authenticity of Brazilian UHT milk: A chemometric approach. **Food Chemistry**, v. 124, p. 692-695, 2011.

STEINSHAMN, H.; INGLINGSTAD, R. A.; EKEBERG, D.; MOLMANN, J.; JORGENSEN, M. Effect of forage type and season on Norwegian dairy goat milk production and quality. **Small Ruminant Research**, v. 122, p. 18-30, 2014.

SUÁREZ, P. L.; SOLDADO, A.; GONZÁLEZ-ARROJO, A.; VICENTE, F.; ROZA-DELGADO, B. Rapid on-site monitoring of fatty acid profile in raw milk using handheld near infrared sensor. **Journal of Food Composition and Analysis**, v. 70, p. 1-8, 2018.

TABET, E.; MANGIA, N. P.; MOUANNES, E.; HASSOUN, G.; HELAL, Z.; DEIANA, P. Characterization of goat milk from Lebanese Baladi breed and his suitability for setting up a ripened cheese using a select starter culture. **Small Ruminant Research**, v. 140, p. 13-17, 2016.

TOMATOKE, H.; OKUYAMA, R.; KATAGIRI, M.; FUZITA, M.; YAMATO, M.; OTA, F. Comparison between Holstein cow's milk and Japanese-Saanem goat's milk in fatty acid composition, lipid digestibility and protein profile. **Bioscience, Biotechnology and Biochemistry**, v. 70, n. 11, p. 2771-2774, 2006.

TRIPATHY, S.; GHOLE, A. R.; DEEP, K.; VANJARI, S. R. K.; SINGH, S. G. A comprehensive approach for milk adulteration detection using inherent bio-physical properties as "universal markers": Towards a miniaturized adulteration detection platform. **Food Chemistry**, v. 217, p. 756-765, 2017.

ULLAH, R.; KHAN, S.; ALI, H.; BILAL, M. Potentiality of using front face fluorescence spectroscopy for quantitative analysis of cow milk adulteration in buffalo milk. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 225, 117518, 2020.

VERRUCK, S.; DANTAS, A.; PRUDENCIO, E. S. Funcionalidade de componentes de leite de cabra, recentes avanços para desenvolvimento de produtos lácteos funcionais e suas implicações na saúde humana. **Journal of Functional Foods**, 52, p. 243-57, 2019.

VIEITEZ, I.; IRIGARAY, B.; CALLEJAS, N.; GONZÁLEZ, V.; GIMENEZ, S.; ARECHAVALETA, A.; GROMOPONE, M.; GÁMBARO, A. Composition of fatty acids and triglycerides in goat cheeses and study of the triglyceride composition of goat

milk and cow milk blends. **Journal of Food Composition and Analysis**, v. 48, p. 95-101, 2016.

WANG, C.; ZHU, Y.; WANG, J. Comparative study on the heat stability of goat milk and cow milk. **Indian Journal of Animal Research**, v. 50, n. 4, p. 610-613, 2016.

WANG, L.; SUN, D. S.; PU, H.; CHENG, J. H. Quality analysis, classification, and authentication of liquid foods by near-infrared spectroscopy: A review of recent research developments. **Critical Reviews in Food Science and Nutrition**, v.57, n. 7, p. 1524-1538, 2017.

WANG, Q.; BOVENHUIS, H. Genome-wide association study for milk infrared wavenumbers. **Journal of Dairy Science**, v. 101, p. 1-13, 2018a.

WANG, K.; YAN, H.; XU, H.; YANG, Q.; ZHANG, S.; PAN, C.; CHEN, H.; ZHU, H.; LIU, J. QU, L.; LAN, X. A novel indel within goat *alpha S1* gene is significantly associated with litter size. **Gene**, v. 671, p. 161-169, 2018b.

WANG, L. L.; LIN, Y. W.; WANG, X. F.; XIAO, N.; XU, Y. D.; LI, H. D.; XU, Q. S. A selective review and comparison for interval variable selection in spectroscopic modeling. **Chemometrics and Intelligent Laboratory Systems**, 172, p. 229-240, 2018c.

WU, D.; HE, Y.; FENG, S. Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment. **Analytica Chimica Acta**, p. 232-242, 2008.

WU, T.; CHEN, H.; LIN, Z.; TAN, C. Identification and quantitation of melamine in milk by near-infrared spectroscopy and chemometrics. **Journal of Spectroscopy**, v. 16, p. 1-8, 2016.

YE, A.; LIU, W.; CUI, J.; KONG, X.; ROY, D.; KONG, Y.; HAN, J.; SINGH, H. Coagulation behavior of milk under gastric digestion: Effect of pasteurization and ultra-high temperature treatment. **Food Chemistry**, v. 286, p. 216-225, 2019.

YUN, Y. H.; LI, H. D.; DENG, B. C.; CAO, D. S. An overview of variable selection methods in multivariate analysis of near-infrared spectra. **Trends in Analytical Chemistry**, v. 113, p. 102-115, 2019.

YURCHENKO, S.; SATS, A.; TATAR, V.; KAART, T.; MOOTSE, H.; JÖUDU, I. Fatty acid profile of milk from Saanen and Swedish Landrace goats. **Food Chemistry**, v. 254, p. 326-332, 2018)

ZHANG, L. G.; ZHANG, X.; NI, L. J.; XUE, Z. B.; GU, X.; HUANG, S. X. Rapid identification of adulterated cow milk by non-linear pattern recognition methods based on near infrared spectroscopy. **Food Chemistry**, v. 145, p. 342-348, 2014.

4 RESULTADOS

A apresentação dos resultados dos três experimentos foi realizada em forma de artigos, de acordo com as normas estabelecidas pelo Programa de Pós-Graduação em Ciência e Tecnologia dos Alimentos (Norma Complementar nº 56/2014).

ARTIGO I- Simultaneous determination of goat milk adulteration with cow milk and their fat and protein contents using NIR spectroscopy and PLS algorithms

ARTIGO II- A low-cost portable NIR spectrophotometer for *in situ* identification of goat milk adulteration with cow milk

ARTIGO III- Imagens digitais como ferramenta para autenticação de leite caprino em termos de adulteração por adição de leite bovino: um estudo de viabilidade

**Simultaneous determination of goat milk adulteration with cow milk and their
fat and protein contents using NIR spectroscopy and PLS algorithms**

Elainy Virginia dos Santos Pereira^{a,b}, David Douglas de Sousa Fernandes^c,
Mário César Ugulino de Araújo^c, Paulo Henrique Gonçalves Dias Diniz^{d*},
Maria Inês Sucupira Maciel^{a,e*}

^a Food Science and Technology Post-Graduate Program - PPGCTA, Technology
Centre, Federal University of Paraíba, Zip Code 58051-900, João Pessoa, Paraíba,
Brazil

^bDepartment of Chemistry - NUPEA, CCT, State University of Paraíba, Zip Code
58429-500, Campina Grande, Paraíba, Brazil

^c Department of Chemistry, CCEN, Federal University of Paraíba, Zip Code 58051-
970, João Pessoa, Paraíba, Brazil

^d Pure and Applied Chemistry Post-Graduate Program, Federal University of West
Bahia, Zip Code 47810-059, Barreiras, Bahia, Brazil

^e Food Science and Technology Post-Graduate Program, Federal Rural University of
Pernambuco, Zip Code 52171-900, Recife, Pernambuco, Brazil

*Corresponding authors:

E-mail addresses: m.inesdcd@gmail.com (M.I.S. Maciel); paulo.diniz@ufob.edu.br
(P.H.G.D. Diniz).

Tel/Fax: +55 81 3320 6536.

26 **ABSTRACT**

27 This work employed NIR spectroscopy and PLS algorithms for the identification
28 and quantification of goat milk adulteration by adding cow milk, besides the
29 determination of their fat and protein contents. Since cow milk can represent a health
30 risk to allergic consumers regardless its amount, PLS-DA was able to identify cow milk
31 additions in goat milk as low as 1.0154 g/100g, likewise the non-adulterated goat and
32 cow milk samples, achieving a 100% of correct classification. For quantification
33 purposes, the Successive Projections Algorithm for interval selection in PLS (*i*SPA-
34 PLS) provided the best results for the determination of both adulteration and fat
35 contents, while PLS gave better results for the protein quantification. Despite the great
36 similarity of both natural dairy matrices and their intrinsic variability, the prediction
37 results provided suitable values with high correlation coefficients and low RMSEP and
38 REP values, with RPD values higher than 3. Therefore, the proposed methodology
39 proved to be a useful, fast and non-destructive tool for screening the quality of goat
40 milk in terms of its adulteration with cow milk, in addition to the quantification of its fat
41 and protein contents.

42

43 **Keywords:** Goat milk; Cow milk; Adulteration; NIR spectroscopy; Successive
44 Projections Algorithm.

45 1. INTRODUCTION

46 World goat milk production surpassed 18 million tonnes per year in 2017,
47 accumulating an increase of approximately 20% in a decade (2007-2017) ([FAOSTAT,](#)
48 [2019](#)). This is due in part to the growing consumer interest in this product and its
49 derivatives, motivated by the unique characteristics of goat milk when compared to cow
50 milk, such as low allergenicity, high digestibility and functional properties ([Hodgkinson](#)
51 [et al., 2012](#); [Ahmed et al., 2015](#)).

52 From the economic point of view, goat milk is more expensive than cow milk and,
53 consequently, its derivatives have a higher added value. Thus, driven by this economic
54 representativeness, adulteration of this food type is a reality. The most fraud is in the
55 replacement category, where there is a partial or total substitution of a food constituent
56 by one or more substitutes (i.e. one adulterant or a mixture of adulterants) of lower cost
57 without the knowledge of consumers. The main adulterant found in goat milk and its
58 derivatives is cow milk, due to its lower cost and greater abundance ([Dabrowska et al.,](#)
59 [2010](#); [Azad, & Ahmed, 2016](#); [Di Pinto et al., 2017](#)).

60 Considering the similarity of these matrices and, consequently, the difficulty of
61 identifying them in mixtures, chromatographic methods, immunological tests,
62 electrophoretic techniques and DNA identification are currently used to determine this
63 kind of adulteration ([Pesic et al., 2011](#); [Song et al., 2011](#); [Kamal, & Karoui, 2015](#); [Di](#)
64 [Pinto et al., 2017](#)). However, such techniques have some drawbacks such as: high cost,
65 time-consuming, laborious and requiring various steps of sample pretreatment.

66 In this scenario, non-destructive methodologies like Near Infrared Spectroscopy
67 (NIRS) have represented an important, interesting and promising alternative for the
68 detection of adulteration in different types of foods, especially for cow milk and its
69 derivatives ([Azad, & Ahmed, 2016](#)). The increase in the popularity of this technique can

70 be attributed to its main characteristics: nondestructive, absence of sample pretreatment,
71 rapidity and accuracy. In comparison with other methodologies based on vibrational
72 spectroscopy (FTIR and RAMAN, for example), the recent advances in the NIR
73 instrumentation, including portable, low-cost and easy-to-use devices, coupled with the
74 construction of the qualitative and quantitative libraries of chemometric models is
75 becoming this technique a reliable, efficient and accessible tool that can be widely
76 employed for different applications in scientific research and industries ([Lohumi, Lee,](#)
77 [Lee & Cho, 2015; Pasquini, 2018](#)).

78 Despite all the methodological advantages, NIR spectroscopy has its limitations,
79 mainly with regard to the sensibility of the technique, which also depends on certain
80 operational and environmental factors, such as the variability of the matrix and the
81 physical state of the sample ([Dupont et al., 2018; Pasquini, 2018](#)). The main sources of
82 variability of milk from various species that affect its chemical composition include
83 race, feed, breeding system (extensive and intensive), lactation period, environmental
84 conditions, age and animal health status ([McDermott et al., 2017; Manousidis et al.,](#)
85 [2018; Miloradovic et al., 2018; Nagy et al., 2019](#)).

86 Classification models based on NIR spectroscopy developed in different studies
87 have already demonstrated the high predictive capacity for the determination of
88 different adulterants in cow milk, and addition of milk from other species in camel milk
89 too ([Chen, Tan, Lin & Wu, 2017; Mabood et al., 2017a; Mabood et al., 2017b;](#)
90 [Karunathilaka et al., 2018](#)). However, to the best of our knowledge, there are no reports
91 on the application of NIR spectroscopy for the determination of goat milk adulteration
92 by adding cow milk.

93 Since in typical NIR spectra exhibit a redundancy of the recorded data due to the
94 strong correlation over the different analytical channels, it has been very useful the use

the noise-reduction properties of Partial Least Squares (PLS) combined with the discard of non-informative or redundant variables. In such scenario, Interval PLS (*i*PLS) and the Successive Projections Algorithm for interval selection in PLS regression (*i*SPA-PLS) have been successfully employed to improve the performance of the regression models, generating more stable models with superior interpretability, and fewer prediction errors. The *i*SPA-PLS algorithm is an extension of SPA that can select multiple intervals of variables, instead the selection of one interval alone as in *i*PLS, which can eventually not contain the required analytical information to solve the problem under study (Gomes, Galvão, Araújo, Vêras, & Silva, 2013; Diniz, Pistonesi, & Araújo, 2015; Krepper et al., 2018). It is important to highlight that NIR and *i*SPA-PLS have already been successfully employed in food analysis for determining protein content in wheat and extract concentration in beer (Gomes et al., 2013), moisture and total polyphenols in tea (Diniz, Pistonesi, & Araújo, 2015), total anthocyanins in jaboticaba fruit (Mariani, Teixeira, Lima, Morgensten, Nardini, & Cunha Júnior, 2015) and fat content in chicken hamburgers (Krepper et al., 2018).

This study is focused on the identification and quantification of goat milk adulteration by adding cow milk, besides the determination of their fat and protein contents employing NIR spectroscopy and PLS algorithms. For classification aims, considering both the economic losses and the potential health risk to allergic consumers regardless of the amount of cow milk ingested, Principal Component Analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA) were tested to differentiate adulterated samples from the pure ones (i.e. non-adulterated goat and cow milk). For quantification, full-spectrum PLS, *i*PLS and *i*SPA-PLS have been employed and then compared.

120 2. MATERIAL AND METHODS

121 2.1. Samples

122 In order to construct a representative model containing different variability
123 sources of the raw materials, 7 lots with 18 samples of goat milk and 7 lots with 18
124 samples of bovine milk were acquired between November 2017 to June 2018 from
125 different farmers located in the micro-regions *Agreste* and *Sertão* of the Paraíba and
126 Pernambuco states, Brazil. The flocks were composed of pure and mestizo animals of
127 different ages with different farrowing number and lactation stages maintained in
128 extensive and semi-extensive breeding systems, as detailed in [Table 1](#).

129 Samples were obtained by accompanying the manual milking of both cows and
130 goats to ensure their purity, i.e. without adulteration. They were then subjected to heat
131 treatment of 65 °C for 30 minutes with subsequent cooling at 4 °C. The samples
132 collected before June 2018 were frozen at -20 °C until the analysis.

133 All 18 samples of goat milk and 18 samples of bovine milk were individually used
134 to construct more representative chemometric models, i.e. they are not mixed to provide
135 a single batch of goat milk and a single batch of bovine milk, respectively. Afterward,
136 goat milk was randomly adulterated by adding cow milk at levels of 1, 3, 5, 7, 10, 20,
137 25, 30, 40, 50 and 75 g/100 g, totalizing 112 mixtures. A total of 20 g for each mixture
138 was prepared in authentic duplicates. The homogenization was performed in Falcon
139 tubes of 50 mL by successive inversions. All NIR measurements were performed in
140 triplicate.

141

142 **Table 1.** Description of the studied goat and cow milk samples.

Lot ^a	Breeding system	Genotype	Period of collection	Farm / Geographical origin	Protein (g/100g)	Fat (g/100g)
Goat 1 (<i>n</i> = 3)	Semi-extensive	Mestizo	Jan/18	Farm 1/ Cabaceiras-PB	3.58±0.03	3.6±0.3
Goat 2 (<i>n</i> = 2)	Semi-extensive	Mestizo	Mar/18	Farm 1/ Cabaceiras-PB	3.64±0.05	2.9±0.1
Goat 3 (<i>n</i> = 3)	Semi-extensive	Mestizo	Apr/18	Farm 1/ Cabaceiras-PB	3.73±0.10	3.5±0.1
Goat 4 (<i>n</i> = 3)	Extensive	Mestizo	Jun/18	Farm 1/ Cabaceiras-PB	3.60±0.23	4.5±0.1
Goat 5 (<i>n</i> = 3)	Extensive	Mestizo	Jun/18	Farm 2/ Cabaceiras-PB	3.55±0.14	3.8±0.1
Goat 6 (<i>n</i> = 2)	Extensive	Mestizo	Jun/18	Farm 3/ Cabaceiras-PB	3.60±0.25	3.4±0.0
Goat 7 (<i>n</i> = 2)	Extensive	Mestizo	Jun/18	Farm 4/ Cabaceiras-PB	3.99±0.04	4.2±0.0
Cow 1 (<i>n</i> = 3)	Semi-extensive	Girolando	Nov/17	Farm 1/ Boa vista-PB	4.54±0.06	1.8±0.1
Cow 2 (<i>n</i> = 3)	Semi-extensive	Girolando	Dec/17	Farm 2/ Cachoeirinha-PE	3.69±0.07	3.0±0.0
Cow 3 (<i>n</i> = 2)	Extensive	Girolando	Jun/18	Farm 3/ Boqueirão-PB	5.03±0.01	1.0±0.0
Cow 4 (<i>n</i> = 3)	Extensive	Mestizo	Jun/18	Farm 4/ Soledade-PB	2.96±0.11	2.0±0.0
Cow 5 (<i>n</i> = 2)	Extensive	Mestizo	Jun/18	Farm 5/ Sumé-PB	4.12±0.23	2.3±0.0
Cow 6 (<i>n</i> = 3)	Extensive	Mestizo	Jun/18	Farm 6/ Serra Branca-PB	3.71±0.15	2.4±0.0
Cow 7 (<i>n</i> = 2)	Extensive	Girolando	Jun/18	Farm 7/ Campina Grande-PB	3.25±0.15	0.9±0.0

143 ^a The number of samples of each lot is indicated into parenthesis.

144 **2.2. Fat and protein determination**

145 The fat content of the samples was determined using the Gerber method
146 (2000.18), according to the recommended by AOAC Official Methods. Protein content
147 was estimated by measuring the nitrogen content of the samples using the Micro-
148 Kjeldhal method and multiplying by the conversion factor (6.38) according to AOAC
149 Official Methods (690.52 and 991.20) ([AOAC, 2005](#)).

150

151 **2.3. Instrumentation and software**

152 **2.3.1 NIR measurements**

153 Samples were disposed into 60 mm × 15 mm polystyrene Petri dishes and then
154 their NIR spectra were acquired in triplicate in the range from 4000 to 10000 cm⁻¹ using
155 an FT-NIR Spectrophotometer, Model Analect Diamond20, coupled with the Diffuse
156 Reflectance Accessory, from Applied Instrument Technologies[®]. Reflectance spectra
157 were recorded using the integrated software, at 8 cm⁻¹ spectral resolution and by
158 integrating 32 scans. All measurements were carried out at room temperature (23 ± 1°C)
159 and used a polytetrafluoroethylene (PTFE) tablet as the blank. An average spectrum
160 from the triplicates of each sample was calculated and then used for the construction of
161 the chemometric models.

162

163 **2.3.2 Chemometric procedure**

164 Initially, as the spectra are noisy and exhibit systematic variations on the baseline,
165 moving mean (MM) to smooth the spectra was always applied before using the
166 following pre-processing techniques: linear baseline correction (LBC), baseline offset
167 (BO), standard normal variate transformation (SNV), and multiplicative scatter
168 correction (MSC).

169 For the classification purposes, PCA and PLS-DA were used as unsupervised and
170 supervised pattern recognition techniques, respectively. For PLS-DA modeling, each
171 class was divided into training (70%) and test (30%) sets by applying the Kennard-
172 Stone (KS) uniform sampling algorithm. PLS-DA was constructed and then validated
173 using a full cross-validation step, while the test samples were only used for the final
174 evaluation of the performance of the model in terms of accuracy, sensitivity and
175 specificity.

176 For multivariate calibration, samples containing from 1.0154 to 100 g/100 g of
177 bovine milk were considered for construction of the calibration model. The non-
178 adulterated cow milk samples (100 g/100 g) were also included in the modeling as
179 representing a possible adulteration by mislabeling. Thus, 90 and 40 samples were
180 selected for the calibration and prediction sets, respectively, by applying the KS
181 algorithm. Then, three different PLS algorithms were used for the construction of the
182 multivariate calibration models: full spectrum PLS, *i*PLS and *i*SPA-PLS. For *i*PLS and
183 *i*SPA-PLS, the spectra were partitioned into 10 and 20 intervals containing the same
184 number of variables. To optimize the number of PLS factors or latent variables (LV) to
185 be included in the calibration set, full cross-validation technique was used, and its
186 efficiency was evaluated according to the lowest root mean square error of cross-
187 validation (RMSECV) and the highest correlation coefficient (r). The prediction set was
188 then only used for the final data evaluation and comparison of the multivariate
189 calibration models. The predictive ability of the final models was evaluated in terms of
190 the lowest values of root mean square error of prediction (RMSEP) and relative error of
191 prediction (REP), highest value of r , and ratio performance deviation (RPD) higher than
192 3. The elliptical joint confidence region (EJCR) test was applied in order to verify the

193 accuracy of the constructed models, in which the results obtained by the chemometric
194 modeling in the prediction set are compared with their respective reference values.

195 Details of the calculations and their respective formulas for all chemometric
196 procedures are well-described in [Fernandes et al. \(2019\)](#). All chemometric procedures
197 used throughout the work were implemented using Matlab® 2019a (Mathworks, USA).

198

199

200 3. RESULTS AND DISCUSSION

201 3.1. Spectra investigation and pre-processing procedures

202 In order to eliminate noise and systematic variations on the baseline, NIR spectra
203 ([Fig. 1](#)) were initially preprocessed with 13-point moving mean and then with different
204 preprocessing methods (SNV; MSC; LBC; LBC-BO; and BO) before the construction
205 of PLS, *i*PLS and *i*SPA-PLS models.

206 As can be seen, a prominent and broad band around 6900 cm^{-1} (the first overtone
207 of O–H stretching vibration) has been attributed to water. The band at 5000 cm^{-1} has
208 been associated with protein as amide, besides absorption bands at about 6250 cm^{-1} (the
209 first harmonic free N–H stretch/amide II combination). Furthermore, other bands
210 corresponding to fat content appeared at 8350 cm^{-1} (the second overtone from –CH
211 stretching), $5800\text{--}5700\text{ cm}^{-1}$ (the first overtone of –CH stretching) and $5300\text{--}5260\text{ cm}^{-1}$
212 (combined –CH stretching and deformation band) ([Nuñez-Sanchez et al., 2016](#); [Dupont
213 et al., 2018](#); [Suárez, Soldado, González-Arrojo, Vicente, & Roza-Delgado, 2018](#)).

214

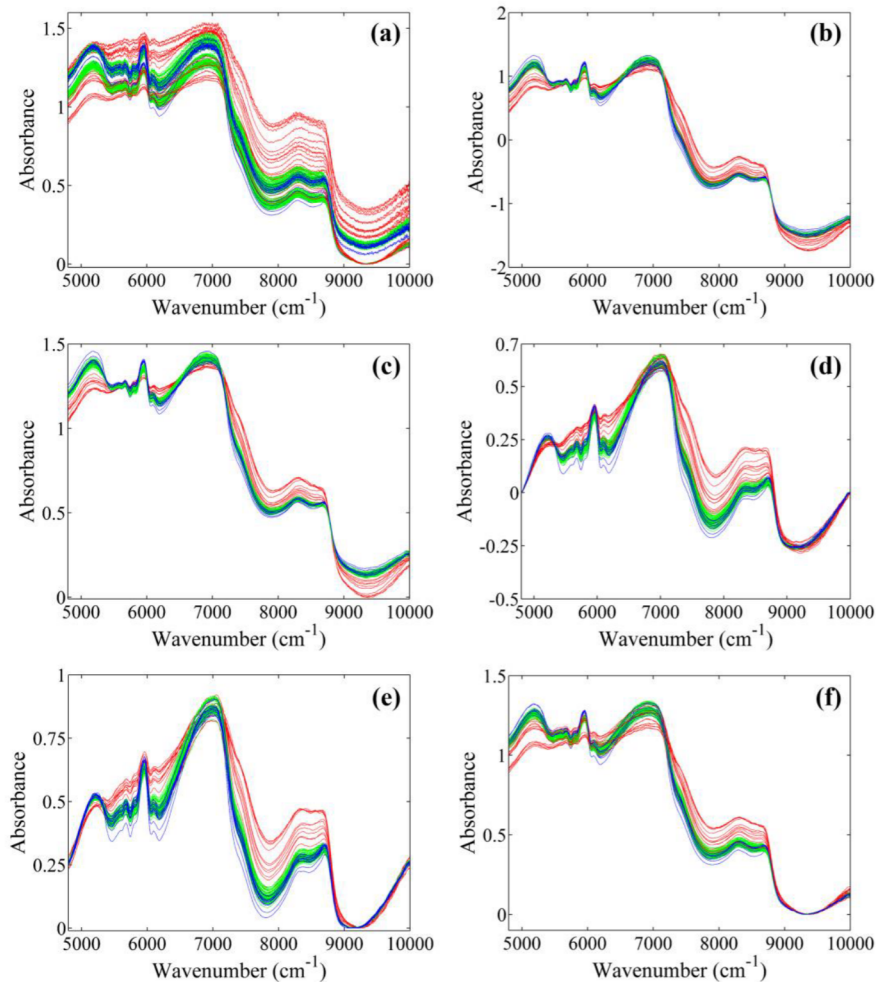


Figure 1. NIR spectra from goat (in blue) and cow (in red) milks and their mixtures (in green): (a) Raw and (b) preprocessed spectra with 13 point moving mean coupled with: SNV, (c) MSC, (d) linear baseline correction, (e) linear baseline correction plus baseline offset and (f) baseline offset.

Additionally, the band related with water content seems very similar in all samples, while goat milk (in blue) presents less variation in the regions attributed to fat and protein content when compared with cow milk (in red). Although this, the spectral profile of all samples was very similar, especially when goat and cow milk are mixed (in green). In fact, the chemical composition of both milk matrices is close, in which the differences are mainly due to the intrinsic variability of the collected samples, as described in [Table 1](#). Consequently, goat milk spectra were close between themselves, because they are produced in a region around 20 km, although the samples were

collected during 6 months from animals of different ages, and number and lactation stages. On the other hand, cow milk samples were collected from animals of different races created in breeding systems in more far regions (around 100 km). Therefore, it can be verified that the factors that influence the milk composition include species, race (genetic similarity) and feed kind (available at a given geographical region), which cause more impact than the seasonality.

This variability between the NIR spectra of milk samples of the same species has been already well described in several studies on authentication, identification of geographical origin and proximate composition of milks from different sources and seasons in order to verify the influence of some factors, such as feed (supplementation and pasture biodiversity), lactation stage and seasonality (Nuñez-Sanchez et al., 2016; Revilla, Escuredo, González-Martín, & Palacios, 2017; Roza-Delgado et al., 2017; Suárez et al., 2018). In this present work, it is worth to note that both the raw material to be adulterated (goat milk) and the adulterant (cow milk) are close natural dairy matrices (see spectra of the non-adulterated samples in blue and red in Fig. 1), which makes its differentiation a challenge.

244

3.2. Identification of adulteration in goat milk with cow milk

Since cow milk can represent economic losses and health risks to allergic consumers, it would be extremely helpful to develop a simple and rapid tool that can identify goat milk adulteration with cow milk regardless of its content. Thus, a non-target analysis employing the NIR spectra as a fingerprint of the samples was performed in order to differentiate non-adulterated goat and cow milk samples as well as adulterated goat milk samples. Thus, an exploratory analysis was initially performed by PCA for screening the natural variability of the samples, whose scores plot is presented

in **Figure 2a**. The five first PCs accounts for more than 99% of the explained variance (see **Figure S1** in the Supplementary Material), being the PC1 and PC2 responsible for 76.96 and 17.65% of the data variance, respectively. However, only PC1 exhibit a trend of separation between both non-adulterated goat (in blue) and cow (in red) milk samples, while the adulterated samples (in green) appear mainly overlapped on goat milk samples. To solve this problem, PLS-DA was applied to the NIR spectral data as a supervised pattern recognition technique and proved to be quite efficient to correctly classify all studied samples into three well-defined classes, as can be clearly seen in **Figure 2b**. As expected, goat milk samples are less dispersed than cow milk samples because of their characteristics discussed previously. Therefore, PLS-DA achieved a 100% of accuracy, sensibility and specificity in the simultaneous identification of non-adulterated goat milk, non-adulterated cow milk and also goat milk adulterated with cow milk. These findings indicate that all chemical information distributed along the whole spectrum allows the recognition of additions as low as 1.0154 g/100g of cow milk added to goat milk when applying PLS-DA.

As fat and protein content are important milk components, they can provoke large overall variance along the whole spectrum; especially fat, because the fat globules and casein micelles in suspension cause high light scattering effect. To investigate this, a biplot of fat and protein contents of all non-adulterated goat and cow milk and also goat milk adulterated with cow milk are present in **Figure 2c**. As can be seen, there are no thresholds that can delimit these three classes, because cow fat content varies from 0.9 to 3.0 g/100g and goat fat content from 2.9 to 4.5 g/100g, while cow protein content ranges from 2.96 to 5.03 g/100g and goat protein content from 3.55 to 3.99 g/100g. Consequently, the adulterated samples have fat and protein contents contained into these ranges, which difficult their separation. These results are in accordance with the

values reported in the literature for goat and cow milks, which presented fat and protein contents of 2.58 to 4.00 g/100 g and 2.38 to 6.36 g/100 g, respectively (Ceballos et al., 2009; Chen et al., 2014; Nunez-Sanchez et al., 2016; Tabet et al., 2016; Roza-Delgado et al., 2017).

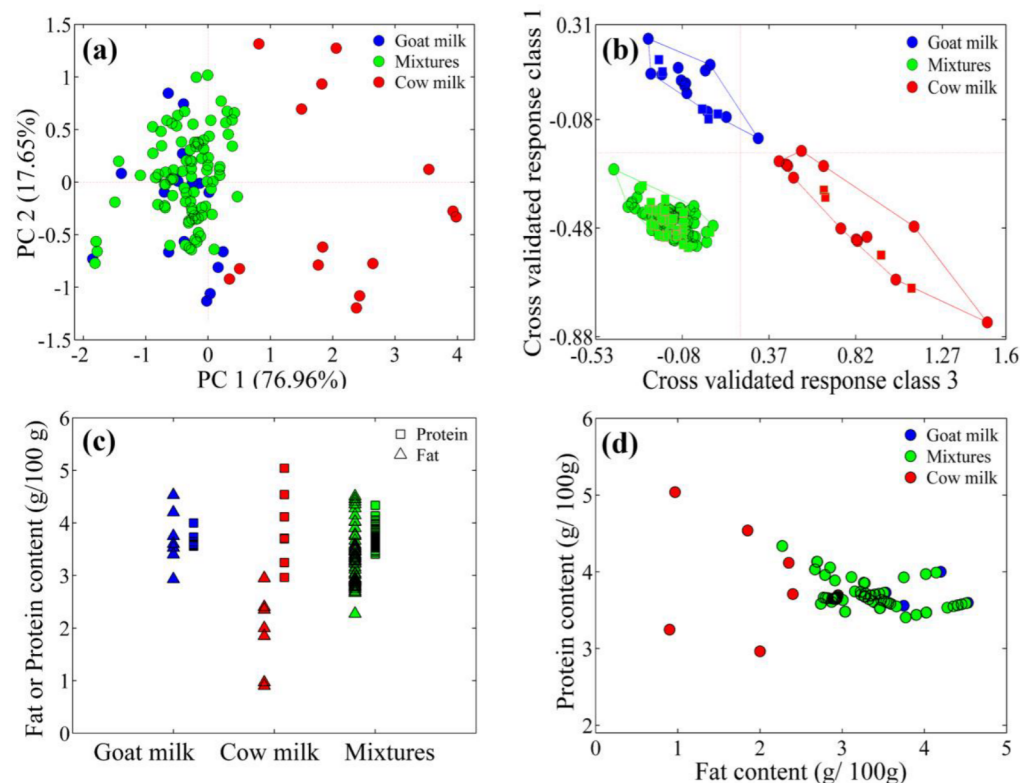


Figure 2. (a) Scores and (b) response plots obtained using the NIR spectra pre-processed with 13MM-BO using PCA and PLS-DA, respectively. (c) Contents and (d) biplot of the fat and protein values obtained by the reference methods for the non-adulterated goat (in blue) and non-adulterated cow milk (in red) samples and their mixtures (in green), respectively.

Although this, some cow milk samples presented abnormal fat content below 2 g/100 g. Different factors can affect the fat content in cow milk, including the nutrition and the milking management. In this sense, milk fat depression (MFD) in lactating cows can occur when the animals are feed with diets supplemented with plants oils, highly fermentable components or with low fiber-high concentrates. As a consequence, milk fat yield is reduced without changes in milk yield or other milk components (Rico &

294 Harvatine, 2013; Leskinen, Ventto, Kairenius, Shingfield & Vilkki, 2019), as observed
 295 here for the results of protein content (**Table 1**). Other factor that can be induce the
 296 abnormal fat content is the incomplete or incorrect milking because the first fractions
 297 before alveolar milk ejection contain low fat content in comparison with the final
 298 fractions (Bansal, Hamann, Grabowski & Singh, 2005; Kahtun, Bruckmaier, Thomson,
 299 House & García, 2019).

300 To confirm that only the information of the fat and protein contents alone are
 301 insufficient to identify the adulteration, a biplot using the fat and protein contents
 302 showed in **Figure 2d** confirms the same behavior of separation indicated in the PCA
 303 scores plot of the **Figure 2a**, i.e. non-adulterated goat milk samples (in blue) are
 304 completely separated from the non-adulterated cow milk samples (in red); however, the
 305 mixtures (in green) are highly overlapped with the non-adulterated goat milk samples.
 306 These corroborates the need of using the entire spectral information to discriminate the
 307 samples, instead of using the fat and protein contents individually.

308

309 **3.3. Determination of adulteration in goat milk with cow milk**

310 The reference values used for the determination of cow milk content used to
 311 adulterate goat milk ranged from 1.0154 to 100 g/100 g, which were then employed for
 312 the construction and prediction steps of the PLS, *i*PLS and *i*SPA-PLS models
 313 employing different pre-processed NIR spectra. The results obtained for the
 314 quantification of the adulteration levels are presented in **Table 2**.

315

316

317

318

Table 2. Results of the determinations of cow milk in goat milk using different pre-processed NIR spectra and PLS algorithms.

	Parameter							
	RMSECV (g/100 g)	r_{CV}	RMSEP (g/100 g)	r_{Pred}	RPD _{Pred}	REP (%)	Bias (t_{cal}) ^b	LV
13 point-moving mean and standard normal variate (13MM-SNV)								
PLS	5.0109	0.9803	9.9654	0.9733	4.00	22.15	0.6661	9
10- <i>i</i> PLS	13.7671	0.8401	21.5224	0.8917	1.85	47.85	2.4253	4
20- <i>i</i> PLS	13.6420	0.8433	14.4887	0.9500	2.75	32.21	2.2818	3
10- <i>i</i> SPA-PLS (9) ^a	10.4393	0.9115	14.5009	0.9561	2.75	32.24	1.4456	5
20- <i>i</i> SPA-PLS (17) ^a	10.4595	0.9111	14.6134	0.9566	2.73	32.49	1.3844	5
13 point-moving mean and multiplicative scatter correction (13MM-MSC)								
PLS	13.6408	0.8296	24.7685	0.8895	1.52	49.48	2.8065	6
10- <i>i</i> PLS	14.7984	0.7956	29.3040	0.7674	1.29	58.54	2.5889	5
20- <i>i</i> PLS	14.9537	0.7906	29.9286	0.7613	1.26	59.79	3.0548	5
10- <i>i</i> SPA-PLS (8) ^a	13.8232	0.8245	24.9205	0.8859	1.52	49.78	2.9030	6
20- <i>i</i> SPA-PLS (10) ^a	13.7008	0.8279	25.0711	0.8853	1.51	50.08	2.8027	6
13 point-moving mean and linear baseline correction (13MM-LBC)								
PLS	1.4276	0.9989	4.1480	0.9936	8.80	14.41	0.4733	10
<i>i</i> PLS	5.7197	0.9820	8.6118	0.9742	4.24	29.92	1.6253	6
20- <i>i</i> PLS	7.9650	0.9648	10.1083	0.9611	3.61	35.12	0.3869	6
10- <i>i</i> SPA-PLS (5) ^a	0.0011	1.0000	6.5763	0.9841	5.55	22.85	0.5663	30
20- <i>i</i> SPA-PLS (14) ^a	1.0862	0.9994	5.0775	0.9905	7.19	17.64	0.3783	12
13 point-moving mean and baseline offset and linear baseline correction (13MM-BO-LBC)								
PLS	0.3851	0.9999	5.6920	0.9908	6.58	17.14	1.4603	16
10- <i>i</i> PLS	22.1480	0.6633	29.2860	0.6423	1.28	88.19	0.6658	2
20- <i>i</i> PLS	7.9827	0.9629	10.0865	0.9632	3.72	30.37	0.3363	5
10- <i>i</i> SPA-PLS (9) ^a	1.1936	0.9990	5.2123	0.9918	7.76	14.97	0.2804	11
20- <i>i</i> SPA-PLS (18) ^a	0.5853	0.9998	5.3209	0.9915	7.61	15.28	0.2247	13
13 point-moving mean and baseline offset (13MM-BO)								
PLS	0.8153	0.9996	3.8197	0.9954	9.86	11.74	0.2001	14
10- <i>i</i> PLS	13.9313	0.8914	16.0572	0.9063	2.35	49.34	0.7952	2
20- <i>i</i> PLS	7.4744	0.9675	8.6889	0.9735	4.34	26.70	0.3571	5
10- <i>i</i> SPA-PLS (9) ^a	1.1936	0.9990	5.2123	0.9918	7.77	14.97	0.2804	11
20-<i>i</i>SPA-PLS (18)^a	0.8016	0.9996	3.6597	0.9955	10.00	11.24	0.4761	14

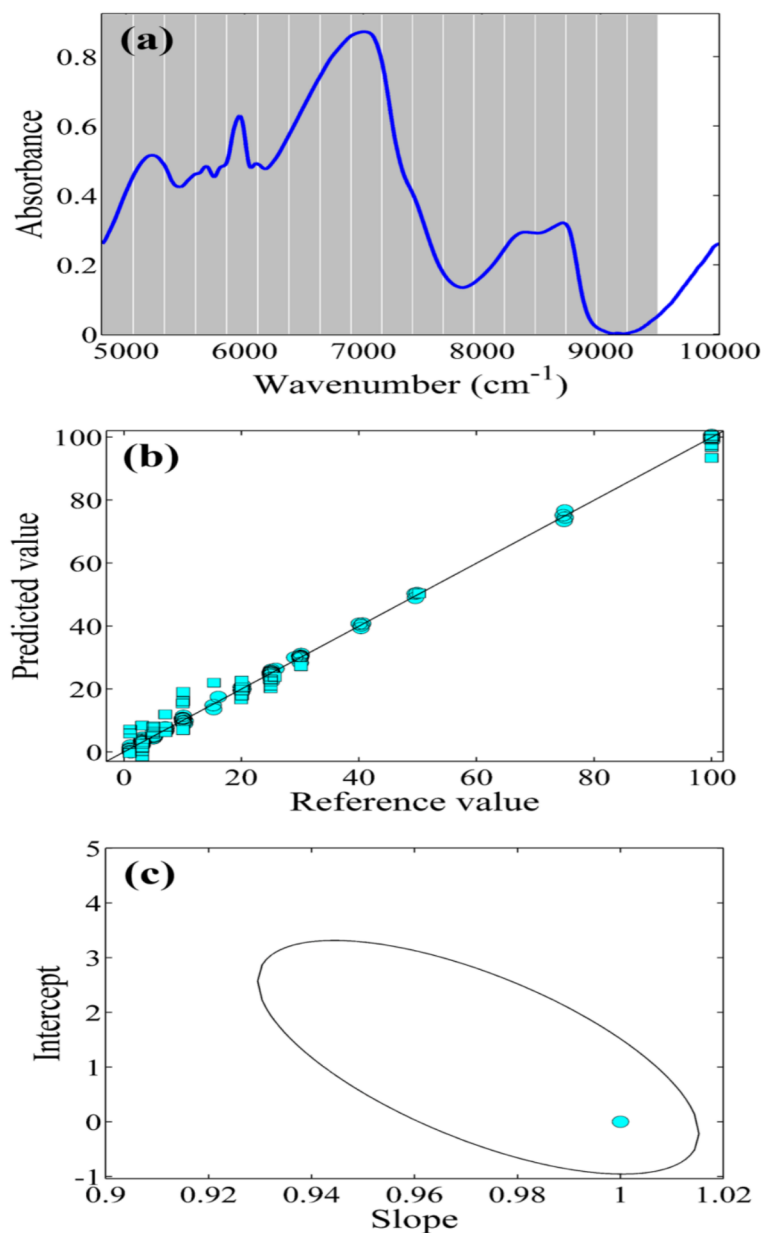
RMSECV: root mean square error of cross-validation; r_{CV} : coefficient of correlation for cross-validation; RMSEP: root mean square error of prediction; r_{Pred} : coefficient of correlation for prediction; RPD_{Pred}: ratio performance to deviation for prediction; t_{calc} : calculated t -value; t_{crit} : critical t -value; LV: latent variables used in the model. REP: relative error of prediction. ^a Number of selected intervals by SPA (e.g. 20-*i*SPA-PLS(18): 18 intervals were selected from 20 by the SPA algorithm); ^b $t_{crit} = 1.6849$.

325

326 As can be seen, the best performance was obtained by using 20-*i*SPA-PLS
 327 modeling for the NIR spectra pre-processed with 13 point-moving mean and baseline
 328 offset (20-*i*SPA-PLS/13MM-BO), reaching r_{CV} of 0.9996 and RMSECV of 0.8016
 329 g/100 g in the calibration set and r_{Pred} of 0.9955, RMSEP of 3.6597 g/100 g, REP of
 330 11.24% and RPD_{Pred} of 10.00 in the prediction set, employing 14 latent variables and
 331 selecting 18 intervals, as indicated in **Figure 3a**. The selection of this large number of
 332 intervals is mainly due to the high complexity and similarity between the goat and cow
 333 milk samples, making it difficult to quantify the addition of one milk matrix into
 334 another. Despite this, the use of interval selection by *i*SPA-PLS eliminates the last two
 335 intervals containing non-informative variables, therefore demonstrating a better
 336 predictive ability to quantify this type of adulteration with an improvement of 0.5% in
 337 the relative prediction error when compared to the *full* PLS modeling coupled with 13-
 338 point moving mean and baseline offset (PLS/13MM-BO). This last model achieved r_{CV}
 339 of 0.9996, RMSECV of 0.8153 g/100 g, r_{Pred} of 0.9954, RMSEP of 3.8197 g/100 g,
 340 REP of 11.74% and RPD_{Pred} of 9.86, also employing 14 latent variables to explain the
 341 variance correlated with the adulteration levels.

342 To illustrate the good fit of the 20-*i*SPA-PLS/13MM-BO model, a predicted vs
 343 reference plot for both calibration and prediction samples is shown in **Figure 3b**.
 344 Additionally, an EJCR test was then applied and demonstrated that the confidence
 345 ellipse of the prediction model includes the theoretical ideal point, presenting, therefore,
 346 no significant bias, as can be seen in **Figure 3c**.

347



348

349 **Figure 3.** (a) Selected intervals and (b) predicted vs reference plot for both calibration (○) and prediction
 350 (□) samples (in cyano) using the 20-*i*SPA-PLS/13MM-BO model. (c) Confidence ellipse of the prediction
 351 model including the theoretical ideal point. The adulterated samples (mixtures) and non-adulterated goat
 352 and cow milk samples are shown in green, blue and red circles, respectively.

353

354 Comparing the results of this present work with those ones from other authors
 355 who investigated the suitability of NIRS for the determination of milk mixtures of
 356 different species (Mabood et al, 2017a; Mabood et al., 2017b), it is possible to verify

that the values obtained for some figures of merit, such as r and RMSE, are relatively similar. However, for these two cases, only three lots of the target milk (camel milk) and three lots of the adulterant milk (cow milk) were used, with additions at merely five adulteration levels (2, 5, 10, 15 and 20 g/100g). In addition, other important merit figures that evidence the good performance of the prediction models such as REP and RPD_{Pred} were not provided. On the other hand, this present work contains more sources of variability of the raw materials and contemplates a wide linear range having eleven different adulteration levels between 1 and 100 g/100g, which can be considered a more representative model.

366

3.4. Determination of fat and protein contents in goat and cow milks and their mixtures

In addition to the determination and identification of adulteration in goat milk by addition of cow milk, NIR spectra were also used to estimate the fat and protein contents of all samples, since they are the components of major interest in the dairy industry, given their different applications. The reference values of the fat and protein contents are illustrated in [Figure 2c](#). Following, pre-processed NIR spectra of the studied samples were then employed for the construction of PLS, i PLS and i SPA-PLS models, as presented in [Tables 3](#) and [4](#).

376

Table 3. Results of the determinations of fat content in cow milk, goat milk and its mixtures using different pre-processed NIR spectra and PLS algorithms.

	Parameter							
	RMSECV (g/100 g)	r_{CV}	RMSEP (g/100 g)	r_{Pred}	RPD _{Pred}	REP (%)	Bias (t_{cal}) ^b	LV
13 point-moving mean and standard normal variate (13MM-SNV)								
PLS	0.22	0.96	0.24	0.92	2.44	7.40	1.48	10
10- <i>i</i> PLS	0.44	0.83	0.44	0.72	1.32	13.66	1.52	5
20- <i>i</i> PLS	0.30	0.93	0.44	0.74	1.31	13.83	0.91	9
10- <i>i</i> SPA-PLS (4) ^a	0.14	0.98	0.24	0.92	2.43	7.45	1.05	13
20- <i>i</i> SPA-PLS (14) ^a	0.10	0.99	0.20	0.94	2.84	6.37	0.90	15
13 point-moving mean and multiplicative scatter correction (13MM-MSC)								
PLS	0.21	0.96	0.24	0.92	2.46	7.34	1.96	11
10- <i>i</i> PLS	0.41	0.85	0.41	0.75	1.43	12.68	1.26	5
20- <i>i</i> PLS	0.31	0.92	0.41	0.75	1.41	12.82	0.78	9
10- <i>i</i> SPA-PLS (5) ^a	0.01	1.00	0.18	0.76	1.37	5.36	1.72	29
20- <i>i</i> SPA-PLS (9) ^a	0.12	0.99	0.23	0.93	2.54	7.11	1.90	14
13 point-moving mean and linear baseline correction (13MM-LBC)								
PLS	0.11	0.98	0.23	0.94	2.96	7.24	1.29	14
<i>i</i> PLS	0.25	0.94	0.45	0.77	1.52	14.12	0.40	8
20- <i>i</i> PLS	0.36	0.88	0.48	0.74	1.42	15.03	0.76	5
10- <i>i</i> SPA-PLS (6) ^a	0.10	0.99	0.22	0.95	3.21	6.68	0.85	14
20-<i>i</i>SPA-PLS (9)^a	0.16	0.98	0.20	0.96	3.36	6.37	0.93	11
13 point-moving mean and baseline offset and linear baseline correction (13MM-BO-LBC)								
PLS	0.17	0.98	0.21	0.94	2.58	6.36	0.14	12
10- <i>i</i> PLS	0.34	0.91	0.36	0.83	1.48	11.08	1.28	7
20- <i>i</i> PLS	0.29	0.93	0.52	0.72	1.03	15.93	1.71	13
10- <i>i</i> SPA-PLS (6) ^a	0.20	0.97	0.23	0.92	2.34	7.02	0.16	11
20- <i>i</i> SPA-PLS (9) ^a	0.16	0.98	0.22	0.92	2.49	6.61	0.04	12
13 point-moving mean and baseline offset (13MM-BO)								
PLS	0.16	0.97	0.32	0.92	2.34	10.00	1.56	12
10- <i>i</i> PLS	0.50	0.73	0.61	0.56	1.21	19.35	0.96	5
20- <i>i</i> PLS	0.60	0.57	0.62	0.57	1.20	19.52	0.21	2
10- <i>i</i> SPA-PLS (6) ^a	0.19	0.96	0.35	0.90	2.13	10.98	1.44	11
20- <i>i</i> SPA-PLS (10) ^a	0.21	0.96	0.35	0.89	2.10	11.13	1.52	10

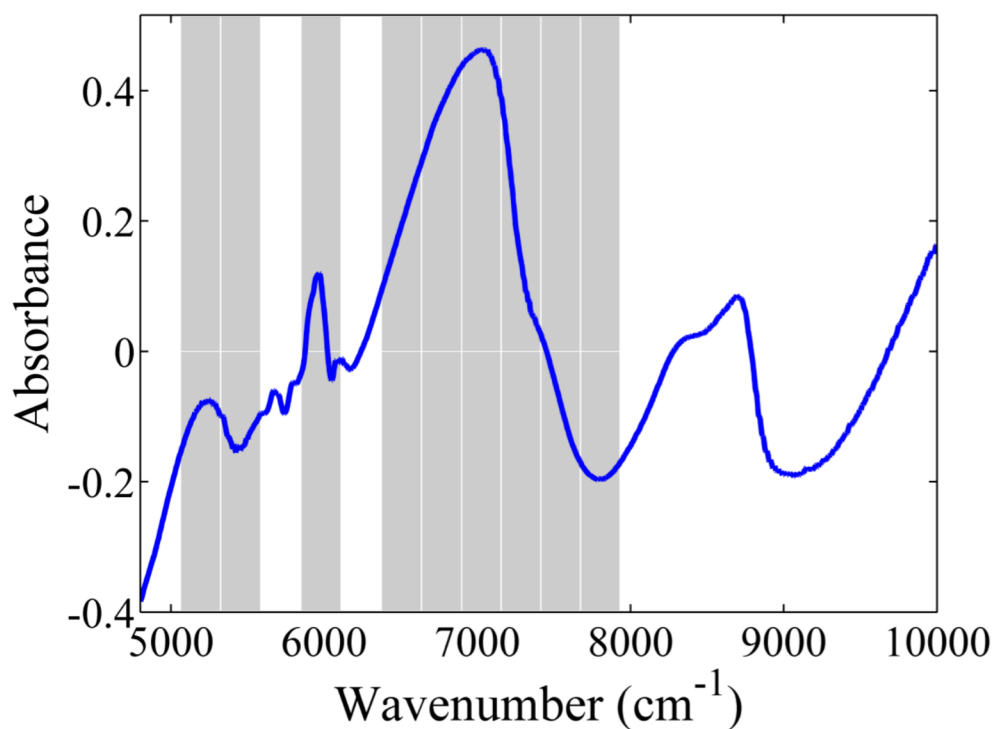
RMSECV: root mean square error of cross-validation; r_{CV} : coefficient of correlation for cross-validation; RMSEP: root mean square error of prediction; r_{Pred} : coefficient of correlation for prediction; RPD_{Pred}: ratio performance to deviation for prediction; t_{calc} : calculated t -value; t_{crit} : critical t -value; LV: latent variables used in the model. REP: relative error of prediction. ^a Number of selected intervals by SPA (e.g. 20-*i*SPA-PLS(18): 18 intervals were selected from 20 by the SPA algorithm); ^b $t_{crit} = 1.6849$.

Table 4. Results of the determinations of protein content in cow milk, goat milk and its mixtures using different pre-processed NIR spectra and PLS algorithms.

	Parameter							
	RMSECV (g/100 g)	r_{CV}	RMSEP (g/100 g)	r_{Pred}	RPD _{Pred}	REP (%)	Bias (t_{cal}) ^b	LV
13 point-moving mean and standard normal variate (13MM-SNV)								
PLS	0.127	0.895	0.166	0.738	1.47	4.46	0.727	7
10- <i>i</i> PLS	0.125	0.899	0.160	0.765	0.53	4.29	1.172	6
20- <i>i</i> PLS	0.118	0.910	0.171	0.724	1.43	4.59	1.087	7
10- <i>i</i> SPA-PLS (8) ^a	0.126	0.897	0.164	0.747	1.49	4.42	0.887	7
20- <i>i</i> SPA-PLS (17) ^a	0.131	0.888	0.170	0.718	1.43	4.58	0.448	7
13 point-moving mean and multiplicative scatter correction (13MM-MSC)								
PLS	0.126	0.897	0.164	0.739	1.48	4.41	0.463	7
10- <i>i</i> PLS	0.152	0.845	0.180	0.691	1.35	4.28	1.170	5
20- <i>i</i> PLS	0.118	0.910	0.172	0.712	1.42	4.62	0.796	7
10- <i>i</i> SPA-PLS (6) ^a	0.104	0.930	0.144	0.821	1.69	3.87	1.696	8
20- <i>i</i> SPA-PLS (12) ^a	0.132	0.886	0.175	0.707	1.39	4.70	0.886	6
13 point-moving mean and linear baseline correction (13MM-LBC)								
PLS	0.039	0.991	0.108	0.907	2.24	2.93	0.824	13
<i>i</i> PLS	0.062	0.976	0.144	0.809	1.69	3.88	0.686	10
20- <i>i</i> PLS	0.174	0.794	0.209	0.530	1.16	5.64	0.133	4
10- <i>i</i> SPA-PLS (7) ^a	0.050	0.985	0.110	0.896	2.21	2.96	0.528	12
20- <i>i</i> SPA-PLS (4) ^a	0.022	0.997	0.142	0.819	1.71	3.83	0.225	15
13 point-moving mean and baseline offset and linear baseline correction (13MM-BO-LBC)								
PLS	0.047	0.989	0.047	0.960	3.54	1.27	1.140	13
10- <i>i</i> PLS	0.121	0.922	0.116	0.813	1.41	3.11	0.050	7
20- <i>i</i> PLS	0.119	0.925	0.108	0.858	1.51	2.90	0.543	8
10- <i>i</i> SPA-PLS (5) ^a	0.006	0.999	0.061	0.930	2.67	1.64	0.445	22
20- <i>i</i> SPA-PLS (10) ^a	0.035	0.994	0.065	0.924	2.49	1.76	0.970	14
13 point-moving mean and baseline offset (13MM-BO)								
PLS	0.130	0.889	0.130	0.854	1.87	3.52	0.565	8
10- <i>i</i> PLS	0.126	0.898	0.167	0.747	1.45	4.53	0.940	6
20- <i>i</i> PLS	0.130	0.889	0.130	0.854	1.87	3.52	0.565	8
10- <i>i</i> SPA-PLS (4) ^a	0.044	0.988	0.133	0.881	1.83	3.59	0.977	12
20- <i>i</i> SPA-PLS (6) ^a	0.147	0.857	0.176	0.695	1.38	4.76	0.525	6

RMSECV: root mean square error of cross-validation; r_{CV} : coefficient of correlation for cross-validation; RMSEP: root mean square error of prediction; r_{Pred} : coefficient of correlation for prediction; RPD_{Pred}: ratio performance to deviation for prediction; t_{calc} : calculated t -value; t_{crit} : critical t -value; LV: latent variables used in the model. REP: relative error of prediction. ^a Number of selected intervals by SPA (e.g. 20-*i*SPA-PLS(18): 18 intervals were selected from 20 by the SPA algorithm); ^b $t_{crit} = 1.6849$.

389 The best performance for fat content determination (**Table 3**) was obtained by
 390 using 20-*i*SPA-PLS for the NIR spectra pre-processed with 13-point moving mean and
 391 linear baseline correction (20-*i*SPA-PLS/13MM-LBC), achieving r_{CV} of 0.98,
 392 RMSECV of 0.16 g/100 g, r_{Pred} of 0.96, RMSEP of 0.20 g/100 g, REP of 6.37% and
 393 RPD_{Pred} of 3.36, employing 11 latent variables to explain the variance among the
 394 samples related to the fat content and selecting only 9 from the 20 possible intervals
 395 indicated in **Figure 4**. As discussed before, these selected intervals are directly related
 396 to water, protein and fat contents, which are correlated due to its centesimal
 397 composition.



398
 399 **Figure 4.** Selected intervals by 20-*i*SPA-PLS/13MM-LBC for construction of the calibration model for
 400 fat content determination.

401 For protein content (**Table 4**), the best performance was achieved by using PLS
 402 for the NIR spectra preprocessed with 13-point moving mean, baseline offset and linear
 403 baseline correction (PLS/13MM-BO-LBC), achieving r_{CV} of 0.989, RMSECV of 0.047
 404 g/100 g, r_{Pred} of 0.960, RMSEP of 0.047 g/100 g, REP of 1.27% and RPD_{Pred} of 3.54.

405 To demonstrate the good fit of the 20-SPA-PLS/13MM-LBC model for fat
 406 determination and the PLS/13MM-BO-LBC model for protein content, predicted vs
 407 reference plots are shown in **Figures 5a** and **5b**, respectively. EJCRC test was then
 408 applied and demonstrated that the ellipses of the prediction models include the
 409 theoretical ideal point for the fat (**Fig. 5c**) and protein (**Fig. 5d**) determinations,
 410 respectively, and therefore they have no biased results.

411 The results obtained in this work resemble those presented by [Roza-Delgado et al.](#)
 412 ([2017](#)), regarding the models for prediction of fat and protein contents in cow milk
 413 produced in different farms and feeding systems. The values presented by these authors
 414 were: r_{CV} of 0.9803 and 0.8222, RMSECV of 0.3193 and 0.3728 g/100 g, RMSEP of
 415 0.3549 and 0.3521 g/100 g for the fat and protein contents, respectively. However, REP
 416 and RPD values were not shown. Comparatively, in this proposed work the values of
 417 RMSECV and RMSEP obtained were smaller, while the correlation coefficients were
 418 similar. Also, [Nunez-Sanchez et al. \(2016\)](#) and [Suárez et al. \(2018\)](#) used the NIRS to
 419 predict the fat content of goat and cow milks with maximum RPD values of 2.19 and
 420 2.39, respectively, which are lower than the recommended value (higher than 3)
 421 consensually accepted in the literature ([Esbensen, Geladi, & Larsen, 2014](#)).

422

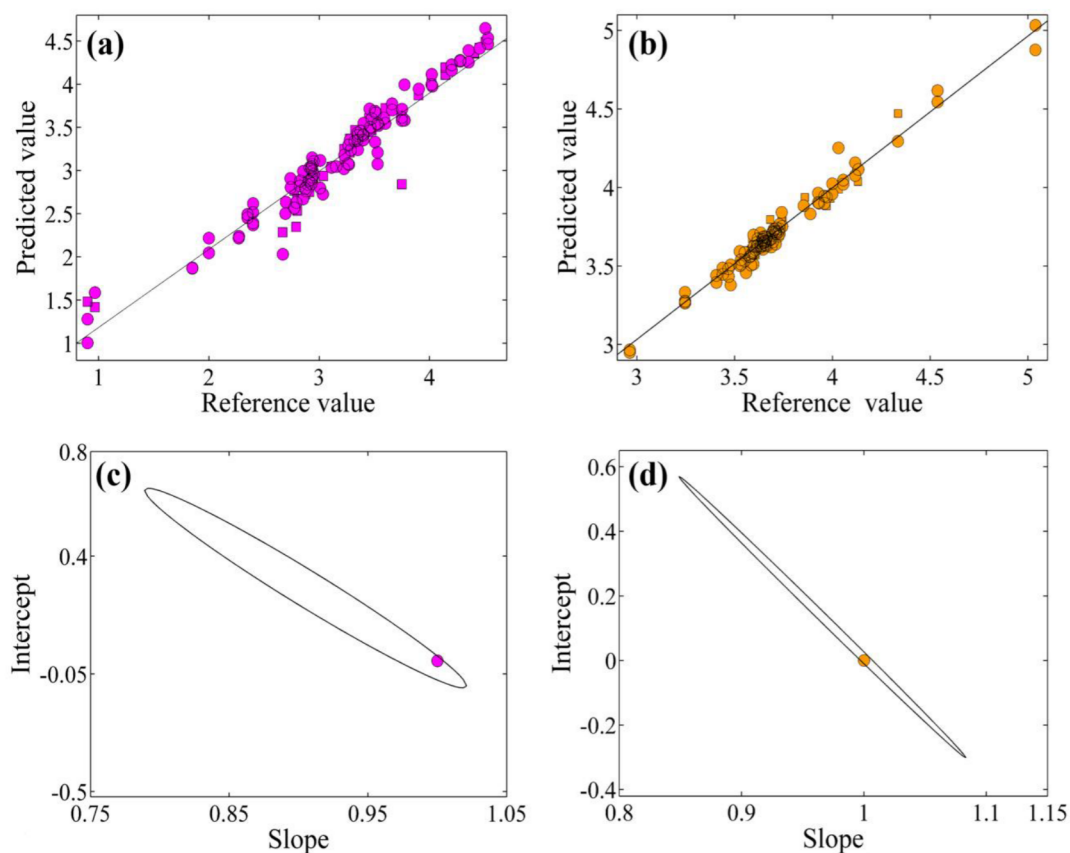


Figure 5. Predicted vs reference plots for both calibration ($\circ$) and prediction ($\square$) samples using (a) the 20-*i*SPA-PLS/13MM-LBC model for fat determination and (b) the PLS/13MM-BO-LBC model for protein content, respectively. Confidence ellipses of the prediction models including the theoretical ideal point for the (c) fat and (d) protein content determinations, respectively.

Considering that the reference methods for fat and protein content determination in milk (Gerber and Kjeldahl methods, respectively) require time, reagents/solvents and pretreatment of the samples, they do not meet the characteristics required for application as platform analyzers. However, according to the results obtained in this study, the models constructed based on the NIRS are able to simultaneously estimate protein and lipid contents in goat and cow milks and their mixtures.

438 **4. Conclusions**

439 In this work, we demonstrate the feasibility of NIR spectroscopy in the
440 identification and quantification of adulteration of goat milk by the addition of cow
441 milk, despite the high complexity in determining one dairy matrix into another. The use
442 of pre-processed NIR spectra coupled with interval selection by *i*SPA-PLS provided the
443 best results for the determination of both adulteration and fat contents, while PLS gave
444 better results for the protein quantification. The good performances obtained in this
445 work can be attributed to the construction of more representative chemometric models
446 that take into account a great sample variability, clearly demonstrated through their
447 figure of merits. Therefore, the proposed methodology proved to be a useful, fast and
448 non-destructive tool for screening the quality of goat milk in terms of its adulteration
449 with cow milk, in addition to its fat and protein contents.

450

451 **ACKNOWLEDGEMENTS**

452 The authors are grateful to Prof. Dr. Mário Araújo for using the NIR spectrometry
453 facilities and the NUPEA-UEPB laboratory for supporting all reference analysis. The
454 authors also thank to CAPRIBOV cooperative and all farmers that provided the milk
455 samples. David Fernandes and Mário Araújo gratefully acknowledge CNPq Brazil for
456 their research fellowships.

457

458 **CONFLICT-OF-INTEREST STATEMENT**

459 The authors declare no competing financial interest. This research did not receive
460 any specific grant from funding agencies in the public, commercial, or not-for-profit
461 sectors.

462

463 **REFERENCES**

- 464 Ahmed, A. S., El-Bassiony, T., Elmalt, L. M., Ibrahim, & H. R. (2015). Identification of
465 potent antioxidant bioactive peptides from goat milk proteins. *Food Research*
466 *International*, 74, 80–88. <https://doi.org/10.1016/j.foodres.2015.04.032>.
- 467 AOAC (2005). Official Methods of Analysis of Association of Official Analytical
468 Chemists. 18th Edition, Washington, DC.
- 469 Azad, T., & Ahmed, S. (2016). Common milk adulteration and their detection
470 techniques. *International Journal of Food Contamination*, 3:22.
471 <https://doi.org/10.1186/s40550-016-0045-3>.
- 472 Bansal, B. K., Hamann, J., Grabowski, N. T., & Singh, K. B. (2005). Variation in the
473 composition of selected milk fraction samples from healthy and mastitic quarters, and
474 its significance for mastitis diagnosis. *Journal of Dairy Research*, 72, 144-152.
475 <https://doi.org/10.1017/S0022029905000798>
- 476 Ceballos, L. S., Morales, E. R., Adarve, G. T., Castro, J. D., Martínez, L. P., &
477 Sampelayo, M. R. S. (2009). Composition of goat and cow milk produced under similar
478 conditions and analyzed by identical methodology. *Journal of Food Composition and*
479 *Analysis*, 22, 322–329. <https://doi.org/10.1016/j.jfca.2008.10.020>.
- 480 Chen, H, Tan, C., Lin, Z. & Wu, T. (2017). Detection of melamine adulteration in milk
481 by near-infrared spectroscopy and one-class partial least squares. *Spectrochimica Acta*
482 *Part A: Molecular and Biomolecular Spectroscopy*, 173, 832-836.
483 <http://dx.doi.org/10.1016/j.saa.2016.10.051>

- 484 Dabrowska, A., Walecka, E., Bania, J. Zelazko, M., Szoltsic, M., & Chrzanowska, J.
 485 (2010). Quality of UHT goat's milk in Poland evaluated by real-time PCR. *Small*
 486 *Ruminant Research*, 94, 32–37. <https://doi.org/10.1016/j.smallrumres.2010.06.005>.
- 487 Di Pinto, A., Terio, V., Marchetti, P., Bottaro, M., Mottola, A., Bozzo, G., Bonerba, E.,
 488 Ceci, E., & Tantillo, G. (2017). DNA-based approach for species identification of goat-
 489 milk products. *Food Chemistry*, 229, 93–97.
 490 <https://doi.org/10.1016/j.foodchem.2017.02.067>.
- 491 Diniz, P. H. G. D., Pistonesi, M. F., & Araújo, M. C. U. (2015). Using iSPA-PLS and
 492 NIR spectroscopy for the determination of total polyphenols and moisture in
 493 commercial tea samples. *Analytical Methods*, 7, 3379–3384.
 494 <https://doi.org/10.1039/C4AY03099K>.
- 495 Dupont, D., Croguennec, T., & Pochet, S. (2018). Milk Proteins – Analytical Methods.
 496 In: *Reference Module in Food Science*, Elsevier. [https://doi.org/10.1016/B978-0-08-](https://doi.org/10.1016/B978-0-08-100596-5.22616-4)
 497 [100596-5.22616-4](https://doi.org/10.1016/B978-0-08-100596-5.22616-4).
- 498 Esbensen, K. H., Geladi, P., & Larsen, A. (2014). The RPD myth. *Mythbusters in*
 499 *Chemometrics*, 25, 5, 24-28. <https://doi.org/10.1255/nirn.1462>.
- 500 FAOSTAT. (2018). Food and Agriculture Organization of the United Nations Statistics
 501 Division. <http://faostat3.fao.org/browse/Q/QL/E>. Accessed 12 January 2019.
- 502 Fernandes, D. D. S., Romeo, F., Krepper, G., Di Nezio, M. S., Pistonesi, M. F.,
 503 Centurión, M. E., Araújo, M. C. U., & Diniz, P. H. G. D. (2019). Quantification and
 504 identification of adulteration in the fat content of chicken hamburgers using digital
 505 images and chemometric tools. *LWT - Food Science and Technology*, 100, 20–27.
 506 <https://doi.org/10.1016/j.lwt.2018.10.034>.

- 507 Gomes, A. A., Galvão, R. K. H., Araújo, M. C. U., Véras, G., & Silva, E. C. (2013).
508 The successive projections algorithm for interval selection in PLS. *Microchemical*
509 *Journal*, 110, 202–208. <https://doi.org/10.1016/j.microc.2013.03.015>.
- 510 Hodgkinson, A. J., McDonald, N. A., Kivits, L. J., Hurford, D. R., Fahey, S., & Prosser,
511 C. (2012). Allergic responses induced by goat milk α s1-casein in a murine model of
512 gastrointestinal atopy. *Journal of Dairy Science*, 95, 1, 83–90.
513 <https://doi.org/10.3168/jds.2011-4829>.
- 514 Kamal, M., & Karoui, R. (2015). Analytical methods coupled with chemometric tools
515 for determining the authenticity and detecting the adulteration of dairy products: A
516 review. *Trends in Food Science & Technology*, 46, 27–48.
517 <https://doi.org/10.1016/j.tifs.2015.07.007>.
- 518 Karunathilaka, S. R., Yakes, B. J.; He, K., Chung, K. K., & Massoba, M. (2018). Non-
519 targeted NIR spectroscopy and SIMCA classification for commercial milk powder
520 authentication: A study using eleven potential adulterants. *Heliyon*, 4, e00806.
521 <https://doi.org/10.1016/j.heliyon.2018.e00806>.
- 522 Khatun, M., Bruckmaier, R. M., Thomson, P. C., House, J., & García, S. C. (2019)
523 Suitability of somatic cell count, electrical conductivity, and lactate dehydrogenase
524 activity in foremilk before versus after alveolar milk ejection for mastitis detection.
525 *Journal of Dairy Science*, 102, 9200–9212. <https://doi.org/10.3168/jds.2018-15752>
- 526 Krepper, G., Romeo, F., Fernandes, D. D. S., Diniz, P. H. G. D., Araújo, M. C. U., Di
527 Nezio, M. S., Pistonesi, M. F., & Centurión, M. E. (2018). Determination of fat content
528 in chicken hamburgers using NIR spectroscopy and the Successive Projections
529 Algorithm for interval selection in PLS regression (iSPA-PLS). *Spectrochimica Acta A*,
530 189, 300–306. <https://doi.org/10.1016/j.saa.2017.08.046>.

- 531 Leskinen, H., Ventto, L., Kairenius, P., Shingfield, K. J., & Vilkki, J. (2019) Temporal
 532 changes in milk fatty acid composition during diet-induced milk fat depression in
 533 lactating cows. *Journal of Dairy Science*, 102, 1-13. [https://doi.org/10.3168/jds.2018-](https://doi.org/10.3168/jds.2018-15860)
 534 [15860](https://doi.org/10.3168/jds.2018-15860)
- 535 Lohumi, S., Lee, S., Lee, H., & Cho, B. K. (2015). A review of vibrational spectroscopy
 536 techniques for the detection of food authenticity and adulteration. *Trends in Food*
 537 *Science & Technology*, 46, 85-98. <http://dx.doi.org/10.1016/j.tifs.2015.08.003>
- 538 Mabood, F., Jabeen, F., Ahmed, M., Hussain, J., Al Mashaykhi, S. A. A. A., Al
 539 Rubaiey, Z. M. A, Farooq, S., Boqué, R., Ali, L., Hussain, Z., Al-Harassi, A., Khan, A.
 540 L., Naureen, Z., Idrees, M., & Manzoor, S. (2017a). Development of new NIR-
 541 spectroscopy method combined with multivariate analysis for detection of adulteration
 542 in camel milk with goat milk. *Food Chemistry*, 221, 746–750.
 543 <https://doi.org/10.1016/j.foodchem.2016.11.109>.
- 544 Mabood, F., Jabeen, F., Hussain, J., Al-Harassi, A., Hamaed, A., Al Mashaykhi, S. A.
 545 A. A., Al Rubaiey, Z. M. A, Manzoor, S., Khan, A., Haq, I. Q. M., Gilani, S. A., &
 546 Khan, A. (2017b). FT-NIRS coupled with chemometric methods as a rapid alternative
 547 tool for the detection & quantification of cow milk adulteration in camel milk samples.
 548 *Vibrational Spectroscopy*, 92, 245–250. <https://doi.org/10.1016/j.vibspec.2017.07.004>.
- 549 Manousidis, T., Parissi, Z. M., Kyriazopoulos, A. P., Malesios, C., Koutroubas, S. D., &
 550 Abas, Z. (2018). Relationships among nutritive value of selected forages, diet
 551 composition and Milk quality in goats grazing in a Mediterranean Woody rangeland.
 552 *Livestock Science*, 218, 8-19. <https://doi.org/10.1016/j.livsci.2018.10.002>.
- 553 Mariani, N. C. T., Teixeira, G. H. A., Lima, K. M. G., Morgenstern, T. B., Nardini, V.,
 554 & Cunha Júnior, L. C. (2015). NIRS and *i*SPA-PLS for predicting total anthocyanin

- 555 content in jaboticaba fruit. *Food Chemistry*, 174, 643–648.
556 <https://doi.org/10.1016/j.foodchem.2014.11.008>.
- 557 McDermott, A., Marchi, M., Berry, M., Visentin, G., Fenelon, M. A., Lopez-Villalobos,
558 N., & McParland, S. (2017). Cow and environmental factors associated with protein
559 fractions and free amino acids predicted using mid-infrared spectroscopy in bovine
560 milk. *Journal of Dairy Science*, 100, 6272–6284. [https://doi.org/10.3168/jds.2016-](https://doi.org/10.3168/jds.2016-12410)
561 [12410](https://doi.org/10.3168/jds.2016-12410).
- 562 Miloradovic, Z., Miocinovic, J., Kljajevic, N., Tomasevic, I., & Pudja, P. (2018). The
563 influence of Milk heat treatment on composition, texture, colour and sensory
564 characteristics of cows' and goats' Quark-type cheeses. *Small Ruminant Research*, 169,
565 154–159. <https://doi.org/10.1016/j.smallrumres.2018.09.012>.
- 566 Nagy, P., Juhász, J., Reiczigel, J., Császár, G., Kocsis, R., & Varga, L. (2019).
567 Circannual changes in major chemical composition of bulk dromedary camel milk as
568 determined by FT-MIR spectroscopy, and factors of variation. *Food Chemistry*, 278,
569 248–253. <https://doi.org/10.3168/jds.2017-12814>.
- 570 Núñez-Sánchez, N., Martínez-Marín, A. L., Polvillo, O., Fernández-Cabanás, V. M.,
571 Carrizosa, J., Urrutia, B., & Seradilla, J. L. (2016). Near Infrared Spectroscopy (NIRS)
572 for the determination of the milk fat fatty acid profile of goats. *Food Chemistry*, 190,
573 244–252. <https://doi.org/10.1016/j.foodchem.2015.05.083>.
- 574 Pasquini, C. (2018). Near Infrared Spectroscopy: A mature analytical technique with
575 new perspectives – A review. *Analytica Chimica Acta*, 1026, 8–36.
576 <https://doi.org/10.1016/j.aca.2018.04.004>.

- 577 Pesic, M., Barac, M., Vrvic, M., Ristic, N., Macej, O., & Stanojevic, S. (2011).
578 Qualitative and quantitative analysis of bovine milk adulteration in caprine and ovine
579 milks using native-PAGE. *Food Chemistry*, 125, 4, 1443–1449.
580 <https://doi.org/10.1016/j.foodchem.2010.10.045>.
- 581 Revilla, I., Escuredo, O., González-Martín, M. I., & Palacios, C. (2017). Fatty acids and
582 fat-soluble vitamins in ewe's milk predicted by near infrared reflectance spectroscopy.
583 Determination of seasonality. *Food Chemistry*, 214, 468–477.
584 <https://doi.org/10.1016/j.foodchem.2016.07.078>.
- 585 Rico, D. E., & Harvatine, K. J. (2013). Induction of and recovery from milk fat
586 depression occurs progressively in dairy cows switched between diets that differ in fiber
587 and oil concentration. *Journal of Dairy Science*, 96, 6621–6630.
588 <http://dx.doi.org/10.3168/jds.2013-6820>
- 589 Roza-Delgado, B., Garrido-Varo, A., Soldado, A., Arrojo, A. G., Valdés, M. C.,
590 Maroto, F., & Pérez-Marín, D. (2017). Matching portable NORS instruments for *in situ*
591 monitoring indicators of milk composition. *Food Control*, 76, 74–81.
592 <https://doi.org/10.1016/j.foodcont.2017.01.004>.
- 593 Song, H., Xue, H., & Han, Y. (2011). Detection of cow's milk in Shaanxi goat's milk
594 with an ELISA assay. *Food Control*, 22, 6, 883–887.
595 <https://doi.org/10.1016/j.foodcont.2010.11.019>.
- 596 Suárez, P. L., Soldado, A., González-Arrojo, A., Vicente, F., & Roza-Delgado, B.
597 (2018). Rapid on-site monitoring of fatty acid profile in raw milk using handheld near
598 infrared sensor. *Journal of Food Composition and Analysis*, 70, 1–8.
599 <https://doi.org/10.1016/j.jfca.2018.03.003>.

- 600 Tabet, E., Mangia, N. P., Mouannes, E., Hassoun, G., Helal, Z., & Deiana, P. (2016).
601 Characterization of goat milk from Lebanese Baladi breed and his suitability for setting
602 up a ripened cheese using a selected starter culture. *Small Ruminant Research*, 140, 13-
603 17. <https://doi.org/10.1016/j.smallrumres.2016.05.011>.

Appendix A. Supplementary Figures

Simultaneous determination of goat milk adulterations with cow milk and their fat and protein contents using NIR spectroscopy and PLS algorithms

Elainy Virginia dos Santos Pereira^{a,b}, David Douglas de Sousa Fernandes^c,
Mário César Ugulino de Araújo^c, Paulo Henrique Gonçalves Dias Diniz^{d,*}, Maria
Inês Sucupira Maciel^{a,e,*}

^a Food Science and Technology Post-Graduate Program - PPGCTA, Technology Centre, Federal University of Paraíba, Zip Code 58051-900, João Pessoa, Paraíba, Brazil

^bDepartment of Chemistry - NUPEA, CCT, State University of Paraíba, Zip Code 58429-500, Campina Grande, Paraíba, Brazil

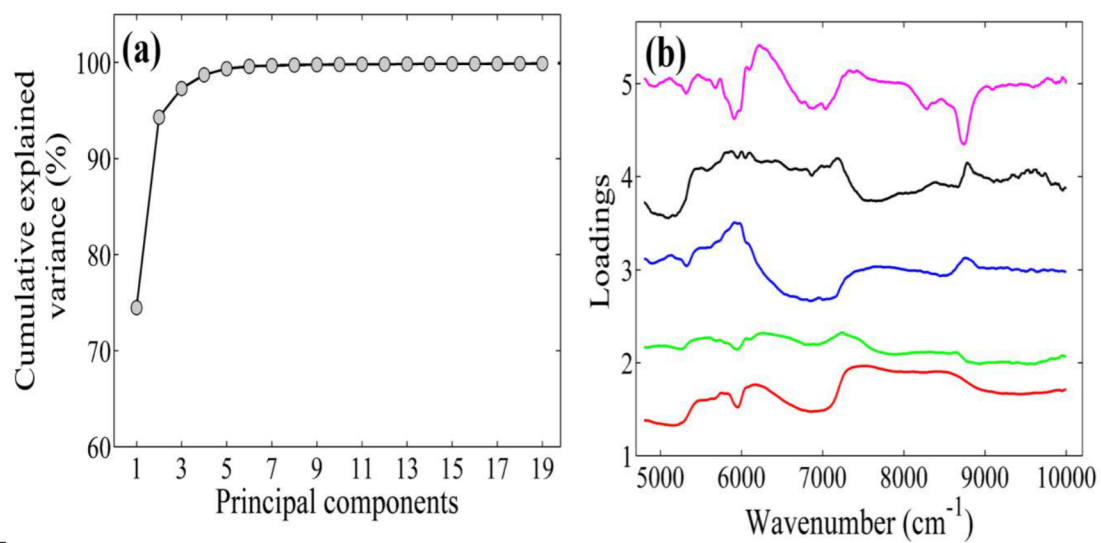
^c Department of Chemistry, CCEN, Federal University of Paraíba, Zip Code 58051-970, João Pessoa, Paraíba, Brazil

^d Pure and Applied Chemistry Post-Graduate Program, Federal University of West Bahia, Zip Code 47810-059, Barreiras, Bahia, Brazil

^e Food Science and Technology Post-Graduate Program, Federal Rural University of Pernambuco, Zip Code 52171-900, Recife, Pernambuco, Brazil

*Corresponding authors:

E-mail addresses: m.inesdcd@gmail.com (M.I.S. Maciel); paulo.diniz@ufob.edu.br (P.H.G.D. Diniz).



■ **Figure S1.** (Supplementary Material). (a) Cumulative explained variance and (b) loadings (for five first PCs) plots obtained by PCA using NIR spectra pre-processed with 13MM-BO.

**A low-cost portable NIR spectrophotometer for the *in situ* identification of goat milk
adulteration with cow milk**

Elainy Virgínia dos Santos Pereira^{a,b}, David Douglas de Sousa Fernandes^c,
Mário César Ugulino de Araújo^c, Paulo Henrique Gonçalves Dias Diniz^{d*}, Maria Inês
Sucupira Maciel^{a,e*}

^a Food Science and Technology Graduate Program - PPGCTA, Technology Centre, Federal
University of Paraíba, Zip Code 58051-900, João Pessoa, Paraíba, Brazil

^b NUPEA, CCT, State University of Paraíba, Zip Code 58429-500, Campina Grande, Paraíba,
Brazil

^c Department of Chemistry, CCEN, Federal University of Paraíba, Zip Code 58051-970, João
Pessoa, Paraíba, Brazil

^d Pure and Applied Chemistry Graduate Program, Federal University of Western Bahia, Zip
Code 47810-059, Barreiras, Bahia, Brazil

^e Food Science and Technology Graduate Program, Federal Rural University of Pernambuco,
Zip Code 52171-900, Recife, Pernambuco, Brazil

*Corresponding authors:

E-mail addresses: m.inesdcd@gmail.com (Maciel, MIS); paulo.diniz@ufob.edu.br,
paulodiniz.quimica@gmail.com (Diniz PHGD).

Tel/Fax: +55 81 3320 6536.

ABSTRACT

This work has investigated the suitability of using a miniaturized near-infrared (NIR) spectrophotometer coupled for the *in situ* identification of goat milk adulteration by adding cow milk. For this, One-Class Partial Least Squares (PLS), PLS for Discriminant Analysis (PLS-DA) and the Successive Projections Algorithm for interval selection in PLS-DA (*iSPA*-PLS-DA) were employed. The best predictive ability was achieved by *iSPA*-PLS-DA when using the spectra pre-processed with smoothing by 7-point window moving mean and baseline offset correction, classifying correctly a 100% of the pure goat milk samples and misclassifying only one adulterated sample in the test set. Therefore, the proposed methodology represents a promise, non-destructive, fast and low-cost tool for screening the authentication of goat milk, also helping to support future decisions by the regulatory agencies in order to prevent this kind of fraud directly in the production local.

Keywords: Milk; Adulteration; Authentication; NIR spectroscopy; Successive Projections Algorithm; Pattern Recognition.

41 1. INTRODUCTION

42 Goat milk consumption has been increased in recent years as an alternative to cow milk
43 due to its low allergenicity, high digestibility and health-promoting compounds associated
44 with important functional properties. These benefits include better calcium absorption,
45 microbiota modulation and anti-atherogenic, antioxidative, anti-platelet, anti-inflammatory
46 and anticarcinogenic properties ([Ahmed, El-Bassiony, Elmalt & Ibrahim, 2015](#); [Clark &](#)
47 [García, 2017](#); [FAOSTAT, 2019](#); [Verruck, Dantas & Prudencio, 2019](#)).

48 In such scenario, the adulteration of commercial goat milk and its derivatives by
49 addition of bovine milk has been a common practice as recently reported in the literature
50 ([Golinelli et al., 2014](#); [Agrimonti, Pirondini, Marmiroli & Marmiroli, 2015](#); [Di Pinto et al.,](#)
51 [2017](#)). From the economic point of view, this type of adulteration is advantageous for
52 producers, since bovine milk has greater abundance and lower added value than goat milk.
53 Moreover, the highly allergenic potential of the bovine milk (particularly among children)
54 when compared with goat milk can represent an important risk for the consumer health ([Clark](#)
55 [& García, 2017](#); [Verruck et al., 2019](#)).

56 Faced with this problem, different researchers have directed their efforts to develop
57 methodologies capable of identifying and/or quantifying this type of adulteration.
58 Chromatographic ([Vieitez et al., 2016](#)), DNA-based ([Golinelli et al., 2014](#); [Di Pinto et al.,](#)
59 [2017](#)), electrophoretic ([Pesic, Barac, Vrvic, Ristic, Macej, & Stanojevic, 2011](#)) and
60 immunoenzymatic ([Song, Xue & Han, 2011](#)) methods demonstrated to be efficient for this
61 objective, although presenting high cost and slow analysis, besides requiring laborious sample
62 pre-treatment steps with high consumption of chemicals and solvents. For this reason, green
63 analytical methodologies based on spectroscopic techniques have been evaluated as
64 alternatives to these traditional methodologies for adulteration detection in different food
65 types, as they have advantages for being non-destructive, of wide applicability and with a

66 minimal sample preparation requirement (Kamal & Karoui, 2015; Nascimento, Santos,
67 Pereira-Filho & Rocha, 2017; Pasquini, 2018).

68 Only recently some authors have demonstrated the suitability of Near-infrared
69 Spectroscopy (NIRS) (Mabood et al. 2017a; Mabood et al. 2017b; Pereira, Fernandes, Araújo,
70 Diniz & Maciel, 2020; Teixeira, Caramês, Baptista, Gigante, & Pallone, 2020) and
71 Synchronous Fluorescence Spectroscopy (Genis, Sezer, Bilge, Durna & Boyaci, 2020; Ullah,
72 Khan, Ali & Bilal, 2020) for the determination of this kind of adulteration by adding milk
73 from one species to another. Despite the numerous advantages, such instruments can only
74 operate in the laboratory, being inadequate for evaluation *in situ*. To overcome this, portable
75 handheld NIR spectrophotometers have recently been employed due to its low cost,
76 portability and robustness. Such instruments have been already successfully employed in
77 determining macroconstituents (such as protein, fat and solids-non-fat) (Muñiz, Cuevas-
78 Valdés & Roza-Delgado, 2020; Roza-Delgado et al., 2017; Suárez, Soldado, González-
79 Arrojo, Vicente & Roza-Delgado, 2018), identification of synthetic adulterants (Henn,
80 Kirchler, Grossgut, & Huck, 2017; Karunathilaka, Yakes, He, Chung & Mossoba, 2018;
81 Santos, Pereira-Filho & Rodriguez-Saona, 2013) and organic milk authentication (Liu, Parra,
82 Pustjens, Hettinga, Mongondry & van Ruth, 2018) in cow milk samples. However, to the best
83 of our knowledge, studies exploiting the use portable miniaturized NIRS instrument for
84 identification of milk mixtures of different species have been not published until now.

85 Given the above, this work investigates the feasibility of using a portable
86 spectrophotometer that provides analytical information in the NIR region to build pattern
87 recognition chemometric models in order to offer a screening tool for the *in situ* identification
88 of goat milk adulteration with cow milk. For this, three different classification approaches
89 based on Partial Least Squares (PLS) have been employed: (i) One-Class PLS (OC-PLS); (ii)

PLS for Discriminant Analysis (PLS-DA); and (iii) the Successive Projections Algorithm for interval selection in PLS-DA (*i*SPA-PLS-DA).

2. MATERIAL AND METHODS

2.1. Samples

In order to construct a representative model containing different variability sources of the studied raw materials, 6 lots of goat milk and 4 lots of bovine milk were acquired in October 2019 from different farmers located in the micro-regions *Agreste*, *Cariri* and *Zona da Mata* of the Paraíba state, Brazil. The goat flocks were composed of animals from the races Alpina Americana, Anglo Nubiana, Moxotó, Murciana and Parda, while the cow flocks were from Sindi, Guzerá and Girolando races. The animals were of different ages with different number and lactation stages maintained in intensive, semi-extensive and extensive breeding systems, as detailed in [Table 1](#).

In order to guarantee the purity (i.e. no adulteration) of the samples, the manual milking of both cows and goats were accompanied *in situ*. Then, they were subjected to heat treatment of 65 °C for 30 minutes with subsequent cooling at 4 °C.

To compose the set of pure goat milk samples, all 6 lots of goat milk were individually and randomly combined, resulting in a class with 54 non-adulterated samples. To construct more representative chemometric models for the class of adulterated samples, all the 6 and 4 lots of goat and cow milk, respectively, were individually used without pooling/mixing them in a single batch of goat milk and a single batch of cow milk. Afterward, goat milk was randomly adulterated by adding cow milk at different levels of 5, 10, 15, 20, 30, 40, 50, 60 and 75 g/100 g, totalizing 146 mixtures. A total of 10 g for each mixture was prepared in authentic duplicates. The homogenization was performed in Falcon tubes of 50 mL by successive inversions. All NIR measurements were performed in triplicate.

115 **Table 1.** Description of the studied goat and cow milk samples.

Lot	Breeding system	Genotype	Sample collection	Farmer	Geographic origin	Fat (g/100 g)	Protein (g/100 g)
Goat 1	Semi-extensive	Moxotó	Oct/19	Farmer 1	Taperoá/Paraíba	5.3±0.1	3.29±0.15
Goat 2	Semi-extensive	Moxotó	Oct/19	Farmer 2	Taperoá/Paraíba	4.6±0.0	3.93±0.13
Goat 3	Semi-extensive	Murciana	Oct/19	Farmer 2	Taperoá/Paraíba	4.6±0.0	3.82±0.08
Goat 4	Semi-extensive	Alpina Americana	Oct/19	Farmer 2	Taperoá/Paraíba	4.2±0.1	3.10±0.23
Goat 5	Semi-extensive	Parda	Oct/19	Farmer 3	Taperoá/Paraíba	4.6±0.0	3.05±0.35
Goat 6	Intensive	Anglo Nubiana	Oct/19	Farmer 4	Campina Grande/Paraíba	5.6±0.0	3.15±0.20
Cow 1	Intensive	Sindi	Oct/19	Farmer 2	Taperoá/Paraíba	4.3±0.0	3.80±0.17
Cow 2	Intensive	Guzerá	Oct/19	Farmer 2	Taperoá/Paraíba	5.1±0.0	3.65±0.10
Cow 3	Semi-extensive	Sindi/Guzerá	Oct/19	Farmer 2	Taperoá/Paraíba	4.1±0.1	3.35±0.25
Cow 4	Extensive	Girolando	Oct/19	Farmer 5	Sobrado/Paraíba	4.6±0.1	3.01±0.15

116

117 **2.2. Fat and protein determination**

118 In order to characterize chemically the main components in the studied milk samples,
 119 their fat and protein contents were determined. The fat content of the samples was determined
 120 using the Gerber method (2000.18), according to the recommended by AOAC Official
 121 Methods. Protein content was estimated by measuring the nitrogen content of the samples
 122 using the Micro-Kjedhal method and multiplying by the conversion factor (6.38) according to
 123 AOAC Official Methods (690.52 and 991.20) (AOAC, 2005).

124

125

126

127 **2.3 Instrumentation and software**

128 **2.3.1 NIR measurements**

129 Samples were disposed into 60 mm × 15 mm polystyrene Petri dishes and their NIR
130 spectra acquired in triplicate in the range from 900 to 1650 nm using a portable NIR
131 instrument, model DLP NIRscan Nano Evaluation Module (Texas Instruments, Dallas, TX).
132 Absorbance spectra were recorded using the integrated software, at 210 digital resolution and
133 by integrating 32 scans with Hadamard transform. All measurements were carried out at room
134 temperature (23 ± 1 °C). An average spectrum from each sample and their triplicates was
135 calculated and then used for the construction of the pattern recognition models.

136

137 **2.3.2 Chemometric procedure**

138 Since the NIR spectra are noisy, moving mean (MM) with 7-point window was initially
139 applied to smooth them. Moreover, they also exhibit systematic variations on the baseline,
140 which required the application of other different pre-processing techniques, such as linear
141 baseline correction (LBC), baseline offset (BO), standard normal variate transformation
142 (SNV), and multiplicative scatter correction (MSC).

143 Considering that the addition of cow milk to goat milk, regardless of the amount, is
144 considered to be adulteration because it can cause economic losses and damage to consumer
145 health, only two classes have been considered for the classification purposes: one class
146 consisting only of 54 pure goat milk samples; and another one containing the 146 adulterated
147 samples.

148 For the construction of the classification models, OC-PLS, PLS-DA and *i*SPA-PLS-DA
149 have been employed and then compared in terms of accuracy, sensitivity, and specificity.
150 Accuracy is calculated as the number of correct classifications divided by the total number of
151 samples in the set under consideration (training or test set). Sensitivity are calculated as the

number of correct positive decisions divided by the total number of known positive cases, while specificity is calculated as the number of correct negative decisions divided by the total number of known negative cases (Massart, Vandeginste, Buydens, Lewi & Smeyers-Verbeke, 1998; Lavine, 2009).

For the OC-PLS approach, forty pure goat milk samples were selected by the Kennard-Stone (KS) uniform sampling algorithm and then used for model construction. The other 14 pure goat milk samples were included together with the 146 adulterated samples in the test set to evaluate the robustness of the model. The confidence levels of the score distance (SD) and the absolute centered residual (ACR) were 0.05. The criterion for choosing the number of latent variables (LVs) was the best compromise between the number of correct answers in both training and test sets.

For the PLS-DA and *i*SPA-PLS-DA approaches, the KS algorithm was used to select 70 and 30% of the samples in each class (i.e. pure and adulterated) for the training and test sets, respectively. Thus, the training and test sets of the pure goat milk class were composed of 38 and 16 samples, respectively, while 102 and 44 mixtures of goat milk with cow milk comprised the training and test sets of the adulterated class, respectively.

All chemometric procedures used throughout the work were implemented using Matlab® 2019a (Mathworks, USA).

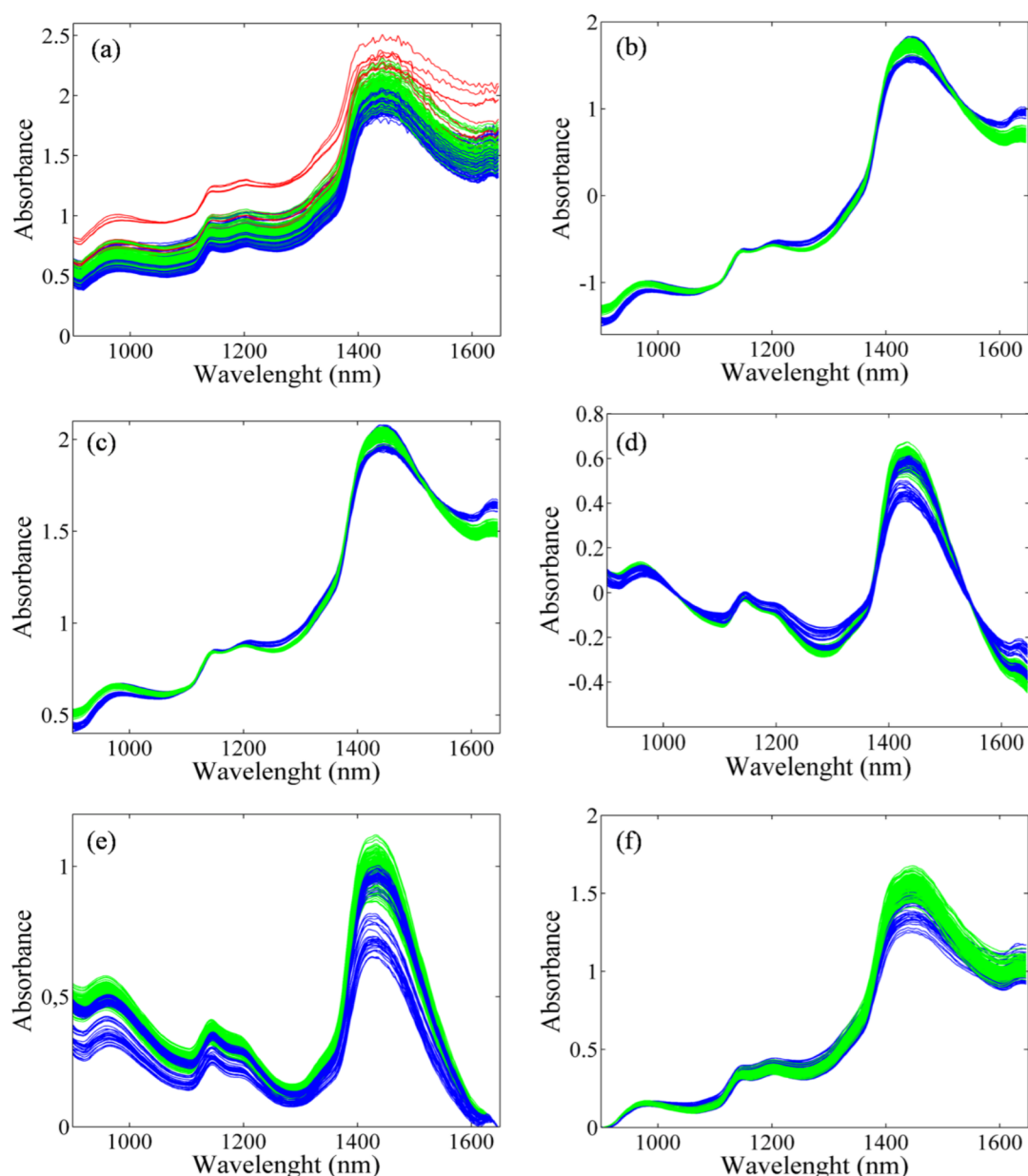
3. RESULTS AND DISCUSSION

3.1. Exploratory analysis

Figure 1a shows the raw NIR spectra of the goat (in blue) and cow (in red) milk samples and their mixtures (in green). As can be seen, the first overtone of O–H stretching vibration of water is responsible for a prominent and broad band around 1450 nm, while small bands between 1100 and 1200 nm correspond to the second overtone from –CH stretch of fat.

177 The first harmonic free N–H stretch/amide II combination accounts for the protein
 178 information around 1630 nm (Nuñez-Sanchez et al., 2016; Dupont, Croguennec & Pochet,
 179 2018; Suárez et al., 2018; Pereira, Fernandes, Araújo, Diniz & Maciel, 2020).

180



181

182 **Figure 1.** NIR spectra from goat (in blue) and cow (in red) milks and their mixtures (in green): raw (a) and pre-
 183 processed spectra with 7-point moving mean coupled with: standard normal variation (b), multiplicative
 184 scattering correction (c), linear baseline correction (d), linear baseline correction plus baseline offset (e) and
 185 baseline offset (f).

Other important observation is the intrinsic high variability of pure goat and cow milk spectra (**Figure 1a**), which are strongly influenced by race, feeding and production system (extensive, semi-extensive and intensive) (Alichanidis, Moatsou & Polychroniadou, 2016; Claeys et al., 2014; Pereira et al., 2020). However, although the sources of variability are the same for both species, their influence on the chemical composition of milk may differ between goats and cows. Chilliard, Toral, Shingfield, Rouel, Leroux and Bernard (2014) concluded that, in general terms, both content and composition of the lipids in goat milk are less influenced by diet than bovine milk. Therefore, this characteristic contributes to the wider dispersion observed for the pure bovine milk spectra when compared to the goat milk spectra. Nevertheless, there is an interspecific similarity between the goat and cow milk samples, which clearly evidenced by the spectra profile as well as the high overlap of the spectra of the mixtures, even when different pre-processing techniques are applied to the spectra (**Figures 1b-1f**). For example, this can be also confirmed by the Principal Component Analysis (PCA) using the 7-point moving window coupled with baseline offset correction (**Figure 2**). Although that the two first PCs account for 96.6% of the explained variance, the PCA scores shows a high overlap between the pure goat milk samples and the other ones adulterated with cow milk (**Figure 2a**). PC1 explains 80.6% of the explained variance and their loadings are mainly influenced by fat and protein, while PC2 describes 16.0% of the explained variance with strong influence of the water band around 1450 nm. Therefore, only by applying the PCA it is not possible to identify this type of adulteration, which requires the use of supervised pattern recognition techniques for this purpose.

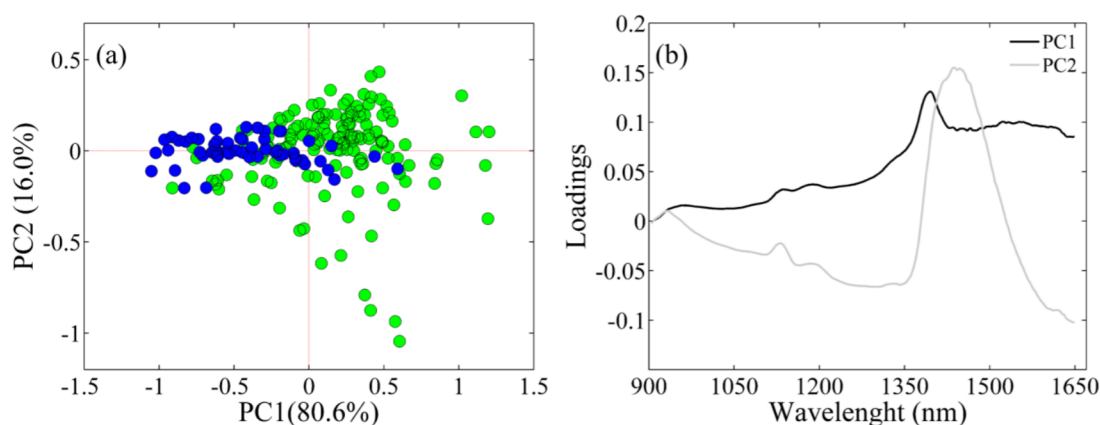


Figure 2. Scores (a) and loadings (b) plots obtained by Principal Component Analysis using the NIR spectra pre-processed with 7-point moving mean and baseline offset correction.

3.2. Identification of goat milk adulteration with cow milk

As the work aims the identification of goat milk adulteration, we evaluated different chemometric strategies depending on the analyst interest. Initially, we tested the use of a one class-modelling strategy because the analyst could be interested in characterizing only the category of pure goat milk. For this, OC-PLS was employed and their results for the different pre-processed NIR spectra are presented in [Table 2](#). As can be seen, all OC-PLS models present high ability to identify correctly the pure goat milk samples with sensitivity above 92.9% using 8 latent variables. In the cases of using the spectra pre-processed with MM+SNV, MM+LBC and MM+BO+LBC, sensitivity achieved a 100% in both training and test sets. However, specificity never exceeded 66.7% regardless of the pre-processing used, which means that goat milk adulterated with cow milk cannot be completely distinguished from pure goat milk due to their high similarity. For illustration, the best general OC-PLS performance was obtained by using the NIR spectra pre-processed with MM+BO, achieving a 100% of correct classification in the training set ([Figure 3a](#)), but misclassifying 1 pure goat milk sample as adulterated and 49 adulterated sample as pure ones in the test set ([Figure 3b](#)).

Table 2 - Classification results of the goat milk adulteration with cow milk using different pre-processed NIR spectra and PLS-based algorithms.

Classifiers	Pre-processing techniques	MM+SNV		MM+MSC		MM+LBC		MM+BO		MM+BO+LBC	
		Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
OCPLS	LVs	8		8		8		8		8	
	Sensitivity	100	100	97.5	100	100	100	100	92.9	100	100
	Specificity	-	41.3	-	50.0	-	55.1	-	66.7	-	59.4
	Accuracy	100	46.7	97.5	54.6	100	59.2	100	69.1	100	63.2
PLS-DA	LVs	17		18		17		16		15	
	Sensitivity	100/100	87.5/93.2	100/100	87.5/97.7	100/100	87.5/86.4	100/97.1	100/84.1	94.7/99.0	81.2/90.9
	Specificity	100/100	93.5/89.5	100/100	97.8/88.2	100/100	87.0/90.9	97.1/100	84.1/100	99.0/94.9	91.5/85.0
	Accuracy	100	91.7	100	95.0	100	86.7	97.9	88.3	97.9	88.3
<i>i</i> 10-SPA-PLS-DA	LVs	10		9		10		10		10	
	Sel. Intervals	7		5		2		4		6	
	Sensitivity	86.8/97.1	87.5/97.7	89.5/97.1	87.5/95.4	89.5/96.1	87.5/79.5	89.5/97.1	100/97.7	94.7/100	81.2/88.6
	Specificity	97.2/87.8	97.8/88.2	97.2/90.2	95.6/88.9	96.2/90.5	80.4/92.0	97.2/90.2	97.7/100	100/94.7	89.4/85.7
	Accuracy	94.3	95.0	95.0	93.3	94.3	81.7	95.0	98.3	98.6	86.7

LVs: latent variables; MM: 7-point moving mean; SNV: standard normal variation; MSC: multiplicative scattering correction; LBC: linear baseline correction; BO: baseline offset.

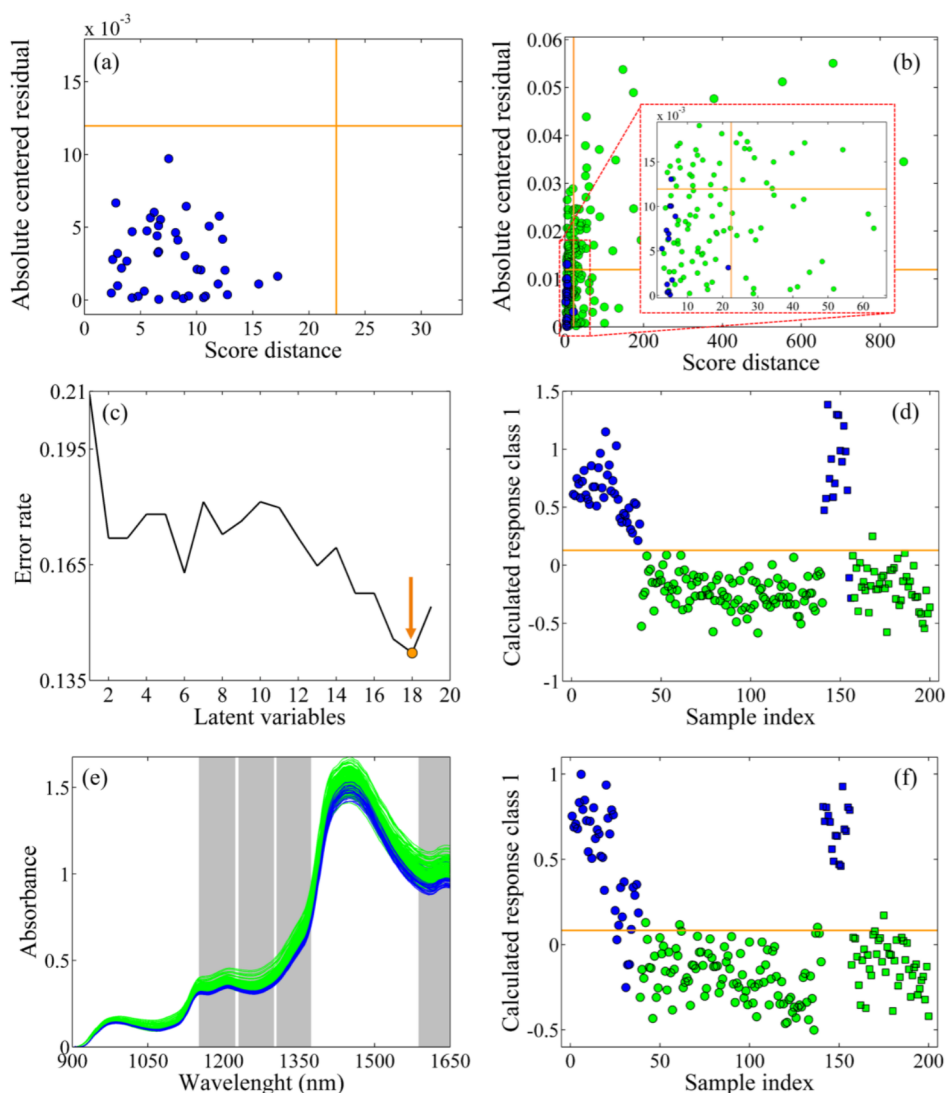


Figure 3. Training (a) and prediction (b) acceptance area for the OC-PLS model using the NIR spectra pre-processed with smoothing by 7-point moving mean and baseline offset correction. Plots of the error rate as a function of the latent variables (c) and its response (d) for the PLS-DA model using the NIR spectra pre-processed with smoothing by 7-point moving mean and multiplicative scattering correction. Plots of the intervals selected by SPA (e) and its response (f) for the 10-*i*SPA-PLS-DA model using the NIR spectra pre-processed with smoothing by 7-point moving mean and baseline offset correction. Orange lines indicate the borders of the area of acceptance in OC-PLS and the thresholds between the studied classes in PLS-DA and 10-*i*SPA-PLS-DA. The pure goat milk samples are represented in blue and the adulterated samples in green.

In fact, in the point of view of the *in situ* identification of this kind of adulteration, the one-class approach should be implemented because only the unadulterated class is the target, even if the adulterant was not known (Oliveri & Downey, 2012; Callao & Ruisánchez, 2018).

246 However, since there is also an appeal by the consumers for health reasons, it is also
247 necessary to model a class of adulterated samples, which requires the use of a two-class
248 approach because the adulterant is known (Callao & Ruisánchez, 2018). Thus, PLS-DA was
249 applied to different pre-processed NIR spectra, whose results are presented in **Table 2**. As can
250 be seen, although the good results for the different pre-processed NIR spectra, the number of
251 latent variables used is very high, indicating a clear overfitting of the constructed PLS-DA
252 models. For illustration, the model PLS-DA/MM+MSC used 18 latent variables (**Figure 3c**)
253 to classify correctly all pure goat milk samples in both training and test sets and misclassify 1
254 adulterated sample as pure goat milk and 2 pure goat milk samples as adulterated in the test
255 set (**Figure 3d**). These results can be directly related to the complexity and similarity of the
256 milk matrices, in addition to the lower resolution of the spectra obtained by the miniaturized
257 NIR spectrometer when compared to the conventional equipment used in the laboratories.
258 These factors increase the correlation of the chemical information over the multiple measured
259 instrumental responses (spectra), which justify the need to use an interval selection technique
260 that can reduce the multicollinearity of the data and, consequently, improve the performance
261 of the models.

262 To circumvent these problems, the noise-reduction properties of PLS-DA for
263 discrimination analysis combined with the discard of non-informative or correlated variables
264 through the interval selection by SPA were also evaluated. The classification results of the
265 *i*SPA-PLS-DA models are presented in **Table 2**. As observed, the best performance was
266 achieved using the NIR spectra pre-processed with MM+BO, employing 10 latent variables
267 and selecting only 4 intervals (**Figure 3e**). In the training set, this model misclassified 4 pure
268 goat milk samples as adulterated and 3 adulterated samples as pure ones, while in the test set
269 only 1 adulterated sample was misclassified as pure and all pure goat milk samples were
270 correctly classified (**Figure 3f**). In other words, the constructed model presents high

271 predictive ability with a 100% of sensitivity, 97,7% of specificity and 98,3% of accuracy in
272 the target class, representing a good alternative to prevent this kind of fraud directly in the
273 production local. This occurs because the 4 intervals selected by the *i*SPA-PLS-DA are
274 directly correlated with the spectral regions associated with the information on fat and protein,
275 as already discussed above (Section 3.1). Additionally, this result corroborates with the
276 evidence that such macromolecules contribute significantly to the differentiation of milk from
277 the goat and bovine species (Mabood et al. 2017a; Mabood et al. 2017b; Pereira, Fernandes,
278 Araújo, Diniz & Maciel, 2020; Genis et al. 2020). Therefore, the selection of intervals by
279 *i*SPA-PLS-DA provides a more parsimonious model, since it is possible to reduce the number
280 of latent variables chosen and select only the most relevant and interpretable chemical
281 information.

282

283 4. CONCLUSION

284 In this work, we demonstrated the suitability of using a miniaturized NIR spectrometer
285 for the *in situ* identification of goat milk adulteration by adding cow milk. To achieve this,
286 different data preprocessing techniques and PLS-based classification algorithms were
287 employed. Among the various models built, the *i*SPA-PLS-DA algorithm coupled with
288 smoothing by 7-point window moving mean and baseline offset correction achieved the best
289 performance in terms of its predictive capacity. As a result, all pure goat milk samples were
290 correctly identified and only one adulterated sample was misclassified in the test set.
291 Therefore, the proposed methodology proved to be a useful, fast and low-cost tool for the
292 non-destructive screening of the authenticity of goat milk in terms of adulteration with cow
293 milk. As a consequence, the proposed methodology can be futurely used as an effective tool
294 to support decisions by the regulatory agencies in order to prevent this kind of fraud directly
295 in the production local.

296

297 **ACKNOWLEDGMENTS**

298 The authors are grateful to Prof. Dr. Mário César Ugulino de Araújo for using the NIR
 299 spectrometry facilities. The authors thanks also to all farmers that provided the milk samples.
 300 David Fernandes and Mário Araújo gratefully acknowledge CNPq Brazil for their research
 301 fellowships.

302

303 **CONFLICT-OF-INTEREST STATEMENT**

304 The authors declare no competing financial interest. This research did not receive any
 305 specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

306

307 **REFERENCES**

- 308 Agrimonti, C., Pirondini, A., Marmiroli, M., & Marmiroli, N. (2015). A quadruplex PCR
 309 (qPCR) assay for adulteration in dairy products. *Food Chemistry*, 187, 58–64.
 310 <https://doi.org/10.1016/j.foodchem.2015.04.017>
- 311 Ahmed, A. S., El-Bassiony, T., Elmalt, L. M., & Ibrahim, H. R. (2015). Identification of
 312 potent antioxidant bioactive peptides from goat milk proteins. *Food Research*
 313 *International*, 74, 80–88. <https://doi.org/10.1016/j.foodres.2015.04.032>
- 314 Alichanidis, E., Moatsou, G., & Polychroniadou, A. (2016). Composition and Properties of
 315 Non-cow Milk and Products. In *Non-Bovine Milk and Milk Products*. Elsevier Inc.
 316 <https://doi.org/10.1016/B978-0-12-803361-6.00005-3>
- 317 AOAC (2005). Official Methods of Analysis of Association of Official Analytical Chemists.
 318 18th Edition, Washington, DC.
- 319 Callao, M. P., & Ruisánchez, I. (2018). An overview of multivariate qualitative methods for
 320 food fraud detection. *Food Control*, 86, 283–293.

- 321 <https://doi.org/10.1016/j.foodcont.2017.11.034>
- 322 Chilliard, Y., Toral, P. G., Shingfield, K. J., Rouel, J., Leroux, C., & Bernard, L. (2014).
 323 Effects of diet and physiological factors on milk fat synthesis, milk fat composition and
 324 lipolysis in the goat: A short review. *Small Ruminant Research*, 122(1–3), 31–37.
 325 <https://doi.org/10.1016/j.smallrumres.2014.07.014>
- 326 Claeys, W. L., Verraes, C., Cardoen, S., De Block, J., Huyghebaert, A., Raes, K., Dewettinck,
 327 K., & Herman, L. (2014). Consumption of raw or heated milk from different species: An
 328 evaluation of the nutritional and potential health benefits. *Food Control*, 42, 188–201.
 329 <https://doi.org/10.1016/j.foodcont.2014.01.045>
- 330 Clark, S., & García, M. B. M. (2017). A 100-Year Review: Advances in goat milk research.
 331 *Journal of Dairy Science*, 100(12), 10026–10044. [https://doi.org/10.3168/jds.2017-](https://doi.org/10.3168/jds.2017-13287)
 332 [13287](https://doi.org/10.3168/jds.2017-13287)
- 333 Di Pinto, A., Terio, V., Marchetti, P., Bottaro, M., Mottola, A., Bozzo, G., Bonerba, E., Ceci,
 334 E., & Tantillo, G. (2017). DNA-based approach for species identification of goat-milk
 335 products. *Food Chemistry*, 229, 93–97. <https://doi.org/10.1016/j.foodchem.2017.02.067>
- 336 Dupont, D., Croguennec, T., & Pochet, S. (2018). Milk Proteins - Analytical Methods. In
 337 *Reference Module in Food Science*. Elsevier. [https://doi.org/10.1016/b978-0-08-100596-](https://doi.org/10.1016/b978-0-08-100596-5.22616-4)
 338 [5.22616-4](https://doi.org/10.1016/b978-0-08-100596-5.22616-4)
- 339 Faostat. (2019). *FAOSTAT*. FAOSTAT Livestock Primary Data.
 340 <http://www.fao.org/faostat/en/#data/QL>
- 341 Genis, D. O., Sezer, B., Bilge, G., Durna, S., & Boyaci, I. H. (2020). Development of
 342 synchronous fluorescence method for identification of cow, goat, ewe and buffalo milk
 343 species. *Food Control*, 108(May 2019). <https://doi.org/10.1016/j.foodcont.2019.106808>
- 344 Golinelli, L. P., Carvalho, A. C., Casaes, R. S., Lopes, C. S. C., Deliza, R., Paschoalin, V. M.
 345 F., & Silva, J. T. (2014). Sensory analysis and species-specific PCR detect bovine milk

- 346 adulteration of frescal (fresh) goat cheese. *Journal of Dairy Science*, 97(11), 6693–6699.
 347 <https://doi.org/10.3168/jds.2014-7990>
- 348 Henn, R., Kirchler, C. G., Grossgut, M. E., & Huck, C. W. (2017). Comparison of sensitivity
 349 to artificial spectral errors and multivariate LOD in NIR spectroscopy – Determining the
 350 performance of miniaturizations on melamine in milk powder. *Talanta*, 166(January),
 351 109–118. <https://doi.org/10.1016/j.talanta.2017.01.035>
- 352 Kamal, M., & Karoui, R. (2015). Analytical methods coupled with chemometric tools for
 353 determining the authenticity and detecting the adulteration of dairy products: A review.
 354 *Trends in Food Science and Technology*, 46(1), 27–48.
 355 <https://doi.org/10.1016/j.tifs.2015.07.007>
- 356 Karunathilaka, S. R., Yakes, B. J., He, K., Chung, J. K., & Mossoba, M. (2018). Non-targeted
 357 NIR spectroscopy and SIMCA classification for commercial milk powder authentication:
 358 A study using eleven potential adulterants. *Heliyon*, 4(9).
 359 <https://doi.org/10.1016/j.heliyon.2018.e00806>
- 360 Lavine, B. K. (2009). Validation of Classifiers. *Comprehensive Chemometrics*, 3, 587–599.
 361 <https://doi.org/10.1016/B978-044452701-1.00027-2>
- 362 Liu, N., Parra, H. A., Pustjens, A., Hettinga, K., Mongondry, P., & van Ruth, S. M. (2018).
 363 Evaluation of portable near-infrared spectroscopy for organic milk authentication.
 364 *Talanta*, 184(December 2017), 128–135. <https://doi.org/10.1016/j.talanta.2018.02.097>
- 365 Mabood, F., Jabeen, F., Ahmed, M., Hussain, J., Al Mashaykhi, S. A. A., Al Rubaiey, Z. M.
 366 A., Farooq, S., Boqué, R., Ali, L., Hussain, Z., Al-Harrasi, A., Khan, A. L., Naureen, Z.,
 367 Idrees, M., & Manzoor, S. (2017a). Development of new NIR-spectroscopy method
 368 combined with multivariate analysis for detection of adulteration in camel milk with goat
 369 milk. *Food Chemistry*, 221, 746–750. <https://doi.org/10.1016/j.foodchem.2016.11.109>
- 370 Mabood, F., Jabeen, F., Hussain, J., Al-Harrasi, A., Hamaed, A., Al Mashaykhi, S. A. A., Al

- 371 Rubaiey, Z. M. A., Manzoor, S., Khan, A., Haq, Q. M. I., Gilani, S. A., & Khan, A.
372 (2017b). FT-NIRS coupled with chemometric methods as a rapid alternative tool for the
373 detection & quantification of cow milk adulteration in camel milk samples. *Vibrational*
374 *Spectroscopy*, 92, 245–250. <https://doi.org/10.1016/j.vibspec.2017.07.004>
- 375 Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., De Jong, S., Lewi, P. J., &
376 Smeyers-Verbeke, J. (1998). *Handbook of Chemometrics and Qualimetrics, Part A*.
377 Elsevier.
- 378 Muñiz, R., Cuevas-Valdés, M., & de la Roza-Delgado, B. (2020). Milk quality control
379 requirement evaluation using a handheld near infrared reflectance spectrophotometer and
380 a bespoke mobile application. *Journal of Food Composition and Analysis*, 86(December
381 2019), 103388. <https://doi.org/10.1016/j.jfca.2019.103388>
- 382 Nascimento, C. F., Santos, P. M., Pereira-Filho, E. R., & Rocha, F. R. P. (2017). Recent
383 advances on determination of milk adulterants. *Food Chemistry*, 221, 1232–1244.
384 <https://doi.org/10.1016/j.foodchem.2016.11.034>
- 385 Núñez-Sánchez, N., Martínez-Marín, A. L., Polvillo, O., Fernández-Cabanás, V. M.,
386 Carrizosa, J., Urrutia, B., & Serradilla, J. M. (2016). Near Infrared Spectroscopy (NIRS)
387 for the determination of the milk fat fatty acid profile of goats. *Food Chemistry*, 190,
388 244–252. <https://doi.org/10.1016/j.foodchem.2015.05.083>
- 389 Oliveri, P., & Downey, G. (2012). Multivariate class modeling for the verification of food-
390 authenticity claims. *TrAC - Trends in Analytical Chemistry*, 35, 74–86.
391 <https://doi.org/10.1016/j.trac.2012.02.005>
- 392 Pasquini, C. (2018). Near infrared spectroscopy: A mature analytical technique with new
393 perspectives – A review. *Analytica Chimica Acta*, 1026, 8–36.
394 <https://doi.org/10.1016/j.aca.2018.04.004>
- 395 Pereira, E. V., Fernandes, D. D. S., Araújo, M. C. U., Diniz, P. H. G. D & Maciel, M. I. S.

- 396 (2020). Simultaneous determination of goat milk adulteration with cow milk and their fat
 397 and protein contents using NIR spectroscopy and PLS algorithms. *LWT- Food Science*
 398 *and Technology*, submitted for publication.
- 399 Pesic, M., Barac, M., Vrvic, M., Ristic, N., Macej, O., & Stanojevic, S. (2011). Qualitative
 400 and quantitative analysis of bovine milk adulteration in caprine and ovine milks using
 401 native-PAGE. *Food Chemistry*, 125(4), 1443–1449.
 402 <https://doi.org/10.1016/j.foodchem.2010.10.045>
- 403 Roza-Delgado, B., Garrido-Varo, A., Soldado, A., González Arrojo, A., Cuevas Valdés, M.,
 404 Maroto, F., & Pérez-Marín, D. (2017). Matching portable NIRS instruments for in situ
 405 monitoring indicators of milk composition. *Food Control*, 76, 74–81.
 406 <https://doi.org/10.1016/j.foodcont.2017.01.004>
- 407 Santos, P. M., Pereira-Filho, E. R., & Rodriguez-Saona, L. E. (2013). Application of hand-
 408 held and portable infrared spectrometers in bovine milk analysis. *Journal of Agricultural*
 409 *and Food Chemistry*, 61(6), 1205–1211. <https://doi.org/10.1021/jf303814g>
- 410 Song, H., Xue, H., & Han, Y. (2011). Detection of cow's milk in Shaanxi goat's milk with an
 411 ELISA assay. *Food Control*, 22(6), 883–887.
 412 <https://doi.org/10.1016/j.foodcont.2010.11.019>
- 413 Suárez, P. L., Soldado, A., González-Arrojo, A., Vicente, F., & de la Roza-Delgado, B.
 414 (2018). Rapid on-site monitoring of fatty acid profile in raw milk using a handheld near
 415 infrared sensor. *Journal of Food Composition and Analysis*, 70(March), 1–8.
 416 <https://doi.org/10.1016/j.jfca.2018.03.003>
- 417 Teixeira, J. L. P., Caramês, E. T. S., Baptista, D. P., Gigante, M. L., & Pallone, J. A. L.
 418 (2020). Vibrational spectroscopy and chemometrics tools for authenticity and
 419 improvement the safety control in goat milk. *Food Control*, 112, 107105.
 420 <https://doi.org/10.1016/j.foodcont.2020.107105>

- 421 Ullah, R., Khan, S., Ali, H., & Bilal, M. (2020). Potentiality of using front face fluorescence
422 spectroscopy for quantitative analysis of cow milk adulteration in buffalo milk.
423 *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 225, 117518.
424 <https://doi.org/10.1016/j.saa.2019.117518>
- 425 Verruck, S., Dantas, A., & Prudencio, E. S. (2019). Functionality of the components from
426 goat's milk, recent advances for functional dairy products development and its
427 implications on human health. *Journal of Functional Foods*, 52, 243–257.
428 <https://doi.org/10.1016/j.jff.2018.11.017>
- 429 Vieitez, I., Irigaray, B., Callejas, N., González, V., Gimenez, S., Arechavaleta, A., Grompone,
430 M., & Gámbaro, A. (2016). Composition of fatty acids and triglycerides in goat cheeses
431 and study of the triglyceride composition of goat milk and cow milk blends. *Journal of*
432 *Food Composition and Analysis*, 48, 95–101. <https://doi.org/10.1016/j.jfca.2016.02.010>

**Imagens digitais como ferramenta para autenticação de leite caprino em termos de
adulteração por adição de leite bovino: um estudo de viabilidade**

Elainy Virginia dos Santos Pereira^{a,b}, David Douglas de Sousa Fernandes^c,
Mário César Ugulino de Araújo^c, Paulo Henrique Gonçalves Dias Diniz^{d*}, Maria Inês
Sucupira Maciel^{e*}

^a Programa de Pós-graduação em Ciência e Tecnologia de Alimentos – PPGCTA Centro de
Tecnologia, Universidade Federal da Paraíba, CEP- 58051-900, João Pessoa, Paraíba,
Brasil

^b Núcleo de Pesquisa e Extensão em Alimentos - NUPEA, Centro de Ciências e Tecnologia,
Universidade Estadual da Paraíba, CEP 58429-500, Campina Grande, Paraíba, Brasil

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

^d Programa de Pós-graduação em Química Pura e Aplicada, Universidade Federal do Oeste
da Bahia, CEP 47810-059, Barreiras, Bahia, Brasil

^e Programa de Pós-graduação em Ciência e Tecnologia de Alimentos, Universidade Federal
Rural de Pernambuco, CEP 52171-900, Recife, Pernambuco, Brasil

23 RESUMO

24 Este trabalho investigou a viabilidade do uso de imagens digitais como ferramenta
25 analítica alternativa e de baixo custo para detecção da adulteração de leite caprino por adição
26 de leite bovino. Para tanto, considerando que este tipo de adulteração, independente do
27 montante, pode acarretar prejuízos econômicos e possíveis danos à saúde do consumidor, os
28 classificadores do tipo “one-class” *One-Class Partial Least Squares* (OCPLS) e *Data Driven-*
29 *Soft Independent Modeling of Class Analogy* (DD-SIMCA) foram empregados como
30 ferramentas de reconhecimento de padrão. Em termos de desempenho de classificação, o
31 melhor resultado foi observado empregando-se DD-SIMCA ao histograma de cor RGB,
32 alcançando 90% de taxa de eficiência na discriminação das amostras. Desta forma,
33 considerando a complexidade da matéria-prima e a acessibilidade da metodologia
34 apresentada, dado seu baixo custo e facilidade de operação, a utilização de imagens digitais
35 como ferramenta para identificação deste tipo de adulteração apresenta potencial para triagem
36 da autenticidade de amostras de leite caprino.

37

38 **Palavras-chave:** Classificadores *One-class*; DD-SIMCA; OC-PLS; Histogramas de cor

39

40 1. INTRODUÇÃO

41 A adulteração em alimentos tem sido um tema frequente e bem relatado nas últimas
42 décadas em publicações científicas. Na maioria dos casos, tal tema tem sido considerado
43 como uma questão econômica, já que frequentemente se dá por substituição e/ou adição de
44 ingredientes de menor valor, e não como um conceito que afeta a saúde pública. Vale
45 considerar, entretanto, que dentre os adulterantes identificados em alimentos, há substâncias
46 tóxicas e alergênicas, que podem acarretar risco direto à saúde do consumidor (MOORE,
47 SPINK e LIPP, 2012; HANDFORD, CAMPBELL e ELLIOTT, 2016; ZHANG e XUE, 2016;
48 NASCIMENTO et al., 2017; KENDALL et al., 2019).

49 Dentre as matrizes alimentícias, o leite caprino figura entre as matérias-primas cuja
50 adulteração tem sido frequentemente relatada (RODRIGUES et al., 2012; GOLINELLI et al.,
51 2014; AGRIMONTI et al., 2015; DI DOMENICO et al., 2017; DI PINTO et al., 2017;
52 CHOOPAN, THANAKIATKRAI e KITPIPIT, 2017; KIM e KIM, 2018). Tal adulteração é
53 particularmente importante, considerando que o principal adulterante detectado é o leite
54 bovino, e que o leite caprino tem seu consumo impulsionado por representar uma alternativa a
55 esta matriz, dadas suas características singulares, tais como baixa alergenicidade, alta
56 digestibilidade e propriedades funcionais (AHMED et al., 2015; CLARK e GARCÍA, 2017;
57 VERRUCK, DANTAS e PRUDENCIO, 2019).

58 Diante desta problemática, diferentes autores têm direcionado seus esforços para
59 desenvolver metodologias capazes de identificar e/ou quantificar esse tipo de adulteração.
60 Métodos cromatográficos (VIEITEZ et al. 2016), baseados em DNA (RODRIGUES et al.,
61 2012; GOLINELLI et al. 2014; DI PINTO et al. 2017; CHOOPAN, THANAKIATKRAI e
62 KITPIPIT, 2017; KIM e KIM, 2018), eletroforéticos (PESIC et al. 2011) e imunoenzimáticos
63 (SONG, XUE e HAN, 2011) foram demonstrados como eficientes para este objetivo. Vale
64 ressaltar, entretanto, que tais metodologias apresentam desvantagens, como elevado custo,

65 lentidão de análise e necessidade de várias etapas de pré-tratamento das amostras, as quais
66 requerem, muitas vezes, grandes quantidades de reagentes e solventes.

67 Neste contexto, metodologias como as baseadas em imagens digitais tem se
68 estabelecido nos últimos anos, tendo em vista o aumento significativo de publicações
69 recentes. A ascensão desta metodologia pode ser atribuída principalmente a duas razões
70 fundamentais. A primeira delas é o elevado potencial de acessibilidade aos analistas,
71 considerando os diversos dispositivos de aquisição de imagens digitais disponíveis. Tais
72 dispositivos têm sido utilizados para obtenção de informação analítica e incluem câmeras
73 digitais (EL-SAYED et al., 2016; FERNANDES et al., 2019), *smartphones* (BOTELHO,
74 DANTAS e SENA, 2017; AZEEM, MADBOULY e EL-SHARAT, 2019), *scanners*
75 (BOTELHO, ASSIS e SENA, 2014; SOUTO et al., 2014) e *webcams* (COSTA et al., 2015;
76 ARAÚJO, MARINHO e GOMES, 2018). Além disso, outra vantagem intrínseca destas
77 ferramentas é o seu enquadramento como metodologia analítica verde, visto que são,
78 geralmente, não destrutivas, não requerem pré-tratamento da amostra, não utilizam solventes
79 e/ou reagentes e não geram resíduos.

80 Para fins analíticos, a informação obtida através da aquisição de imagens digitais é
81 extraída sob a forma de histogramas de cor, que por sua vez descrevem a distribuição da
82 frequência dos pixels em função de cada componente de cor. Considerando que a cor é uma
83 propriedade física aparente, atribuída à distribuição de um espectro de luz, os histogramas
84 dependem da fonte de iluminação, a qual emite a luz que interage com o objeto (amostra) e,
85 por sua vez, é então registrada pelo dispositivo de captura de imagem (DINIZ et al., 2012;
86 McANDREW, 2015; FERNANDES et al., 2019).

87 Dada a natureza multivariada dos dados provenientes dos histogramas de cor, a
88 combinação com ferramentas quimiométricas para acessar as informações químicas e/ou
89 físicas relevantes é imprescindível. No campo das adulterações de alimentos, metodologias

analíticas baseadas em histogramas de cor acoplados a técnicas de reconhecimento de padrões têm sido frequentemente utilizadas quando o objetivo principal é identificar e/ou diferenciar uma ou mais classes de amostras, como no caso de identificação de adulterações (tipificadas ou não) ou rastreamento de origem geográfica de alimentos (DINIZ et al., 2012; SOUTO et al., 2015; ARAÚJO, MARINHO e GOMES, 2018; FERNANDES et al., 2019; LIMA et al., 2020).

Ante o exposto, considerando que a adição de leite de vaca, independente do montante, é uma das principais formas de adulteração do leite de cabra, podendo acarretar prejuízos econômicos e possíveis danos à saúde do consumidor, objetivou-se investigar a viabilidade do uso de imagens digitais como ferramenta analítica alternativa e de baixo custo para detecção deste tipo de adulteração. Para isto, a autenticidade do leite de cabra será verificada através do uso de classificadores “one-class”; neste caso, *One-Class Partial Least Squares* (OCPLS) (XU et al., 2013) e *Data Driven-Soft Independent Modeling of Class Analogy* (DD-SIMCA) (ZONTOV et al., 2017). Nestes tipos de classificadores, apenas a classe alvo (isto é, a classe não adulterada) é modelada e as amostras que não sejam incluídas na classe alvo serão consideradas adulteradas.

106

107 2. MATERIAL E MÉTODOS

108 2.1. Amostras

Com o objetivo de construir um modelo representativo contendo diferentes fontes de variabilidade na matéria-prima estudada, 6 lotes de leite caprino e 4 lotes de leite bovino foram adquiridos em outubro de 2019 de diferentes produtores localizados nas microrregiões do Agreste, Cariri e Zona da Mata da Paraíba, Brasil. De forma que, cada lote representou um rebanho diferente. Os rebanhos caprinos eram constituídos das raças Alpina Americana, Anglo Nubiana, Moxotó, Murciana e Parda, enquanto os rebanhos bovinos eram das raças

Sindi, Guzerá e Girolando. Todos os rebanhos eram constituídos de animais de diferentes idades e períodos de lactação. O sistema de manejo variou entre extensivo, semi-extensivo e intensivo, como detalhado na [Tabela 1](#).

Tabela 1. Descrição das amostras de leite caprino e bovino analisadas.

Lote	Tipo de criação	Raça	Período de coleta	Localização geográfica	Lipídeos* (g/100 g)	Proteínas* (g/100 g)
Caprino 1	Semi-extensivo	Moxotó	Out/19	Taperoá/PB	5.3±0.1	3.29±0.15
Caprino 2	Semi-extensivo	Moxotó	Out/19	Taperoá/PB	4.6±0.0	3.93±0.13
Caprino 3	Semi-extensivo	Murciana	Out/19	Taperoá/PB	4.6±0.0	3.82±0.08
Caprino 4	Semi-extensivo	Alpina Americana	Out/19	Taperoá/PB	4.2±0.1	3.10±0.23
Caprino 5	Semi-extensivo	Parda	Out/19	Taperoá/PB	4.6±0.0	3.05±0.35
Caprino 6	Intensivo	Anglo Nubiana	Out/19	Campina Grande/PB	5.6±0.0	3.15±0.20
Bovino 1	Intensivo	Sindi	Out/19	Taperoá/PB	4.3±0.0	3.80±0.17
Bovino 2	Intensivo	Guzerá	Out/19	Taperoá/PB	5.1±0.0	3.65±0.10
Bovino 3	Semi-extensivo	Sindi/ Guzerá	Out/19	Taperoá/PB	4.1±0.1	3.35±0.25
Bovino 4	Semi-extensivo	Girolando	Out/19	Sobrado/PB	4.6±0.1	3.01±0.15

*As determinações de proteínas e lipídeos foram realizadas segundo as metodologias oficiais (métodos micro-Kjedhal (690.52 e 991.20) e Gerber (2000.18), respectivamente) ([AOAC, 2005](#)).

A fim de garantir a pureza (i.e., ausência de adulteração) das amostras, a ordenha manual das cabras e vacas foram acompanhadas *in situ*. As amostras foram transportadas sob refrigeração e posteriormente pasteurizadas a 65 °C por 30 minutos e resfriadas a 4 °C.

Para compor o conjunto de amostras puras de leite caprino, os 6 lotes foram aleatoriamente combinados, resultando em uma classe com 54 amostras. As amostras adulteradas de leite caprino foram obtidas mediante mistura individual e aleatória dos diferentes lotes de leites puros caprino e bovino nos níveis de 5, 10, 15, 20, 30, 40, 50, 60 e 75 g/100 g, totalizando 138 misturas. Um total de 10 g para cada mistura foi preparado em

131 duplicatas autênticas. A homogeneização foi realizada em tubos Falcon de 50 mL por
132 inversões sucessivas. As amostras foram dispostas em placas de Petri lisas de poliestireno (60
133 × 15 mm) para posterior aquisição imagens digitais.

134

135 **2.2. Aquisição das Imagens Digitais**

136 As imagens foram obtidas utilizando o scanner de uma impressora multifuncional HP,
137 modelo Deskjet 4180 com 300 dpi de resolução, gerando uma matriz de dimensões 2550 x
138 3507 pixels. A imagem de cada amostra foi adquirida individualmente, em uma mesma área
139 do scanner, para padronizar os efeitos da incidência da luz. As imagens foram obtidas em
140 triplicata, a partir das quais foram obtidos histogramas médios de cada amostra a partir de
141 uma região de interesse (ROI) circular localizada no centro de cada imagem. Os histogramas
142 nos sistemas de cor escala de cinzas (GS), vermelho, verde e azul (RGB) e matiz, saturação e
143 intensidade (HSI) foram obtidos a partir de cada imagem digital usando a interface gráfica
144 Image_GUI, escrita em ambiente Matlab e disponibilizada gratuitamente em:
145 http://laqa.quimica.ufpb.br/Softwares%20para%20Download/Imagens_gui.rar.

146 Considerando que cada componente de cor é representado digitalmente por 8 bits,
147 assumindo valores entre 0 e 255 (totalizando 256 variáveis), um vetor **x** de informação
148 instrumental com dimensões de 1×1792 contemplando todos os componentes de cor avaliados
149 neste trabalho foi obtido para cada amostra. Os histogramas de cor foram avaliados
150 individualmente (GS, RGB e HSI) e em combinação (GS+RGB, GS+HSI, RGB+HSI e
151 GS+RGB+HSI).

152

153 **2.3. Análise quimiométrica dos dados**

154 Tendo em vista que a estratégia a ser avaliada neste trabalho consiste em empregar
155 técnicas de reconhecimento de padrão em que apenas a classe alvo é modelada,

assemelhando-se a situações reais, nas quais não é possível caracterizar a classe de amostras adulteradas, apenas duas classes foram consideradas para fins de classificação. A classe alvo foi composta por 54 amostras puras de leite caprino e a outra classe por 138 amostras adulteradas (i.e., misturas de leite caprino e bovino).

Para construção dos modelos OC-PLS e DD-SIMCA, 40 amostras de leite caprino puro foram selecionadas pelo algoritmo de amostragem uniforme Kennard-Stone (KS) como conjunto de treinamento. As 14 amostras de leite caprino puro restantes foram incluídas no grupo de teste com as 138 amostras adulteradas, para avaliar a robustez dos modelos construídos.

As performances dos modelos foram avaliadas em termos de sensibilidade, especificidade e eficiência. A sensibilidade foi calculada como o número de amostras corretamente classificadas na classe dividido pelo número total de amostras pertencentes àquela classe, enquanto especificidade foi calculada como o número de decisões negativas corretas (negativos verdadeiros) dividido pelo total de amostras negativas (negativos verdadeiros e falso positivas). Já a eficiência foi calculada como a média geométrica da sensibilidade e da especificidade (MASSART et al., 1998; LAVINE, 2009; OLIVERI e DOWNEY, 2012).

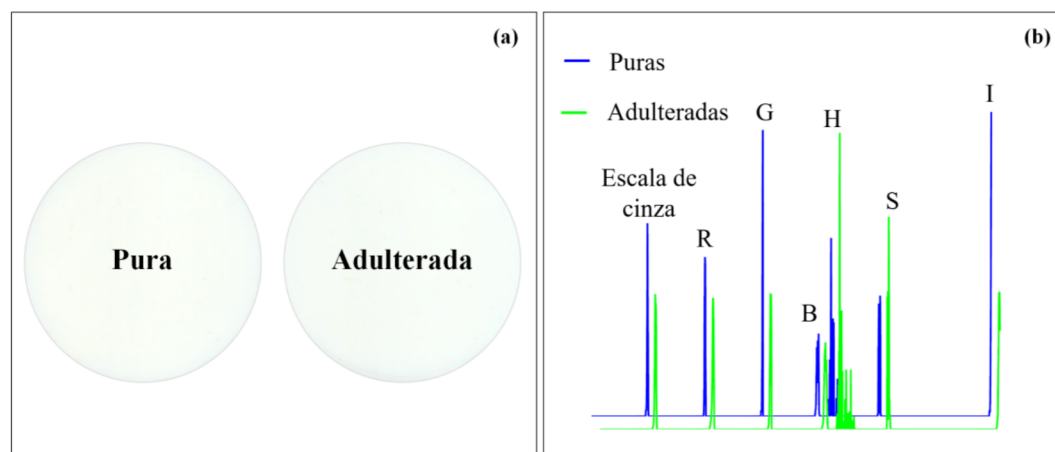
Para OC-PLS, os níveis de confiança para distância de escore (SD) e para resíduo absoluto centralizado (ACR) foram de 0,05. Em DD-SIMCA, o nível de significância para detecção de *outlier* foi de 0,01, a área de aceitação foi baseada na distribuição qui-quadrado no modo robusto. O número apropriado de variáveis latentes (LVs) e componentes principais (PCs) para OC-PLS e DD-SIMCA, respectivamente, foi obtido examinando os valores de sensibilidade do modelo no conjunto de treinamento.

Todos os procedimentos quimiométricos foram implementados em Matlab[®] 2019a (Mathworks, USA).

181 3. RESULTADOS E DISCUSSÃO

182 Os histogramas médios das amostras puras (azuis) e adulteradas (verdes) estão
 183 representados na **Figura 1**. É possível evidenciar que os histogramas de ambas as classes são
 184 bem similares e sobrepostos, o que justifica a utilização de ferramentas quimiométricas, tendo
 185 em vista que uma simples inspeção visual não permite separação das classes. Além disso,
 186 considerando que as informações contidas nos diferentes canais analíticos observados estão
 187 correlacionadas (McANDREW, 2015), fica evidente a necessidade de explorar toda a
 188 informação registrada, pois as possíveis fontes de diferenciação entre as classes podem estar
 189 distribuídas ao longo dos canais registrados.

190



191 **Figura 1.** (a) Ilustrações da mesma amostra de leite caprino sem e com a adição de leite bovino. (b) Histogramas
 192 de cor médios obtidos para as amostras puras (linha azul) e adulteradas (linha verde) de leite caprino nos
 193 diferentes canais analíticos: R – vermelho, G – verde, B – azul, H – matiz, S – saturação, I – intensidade.

195

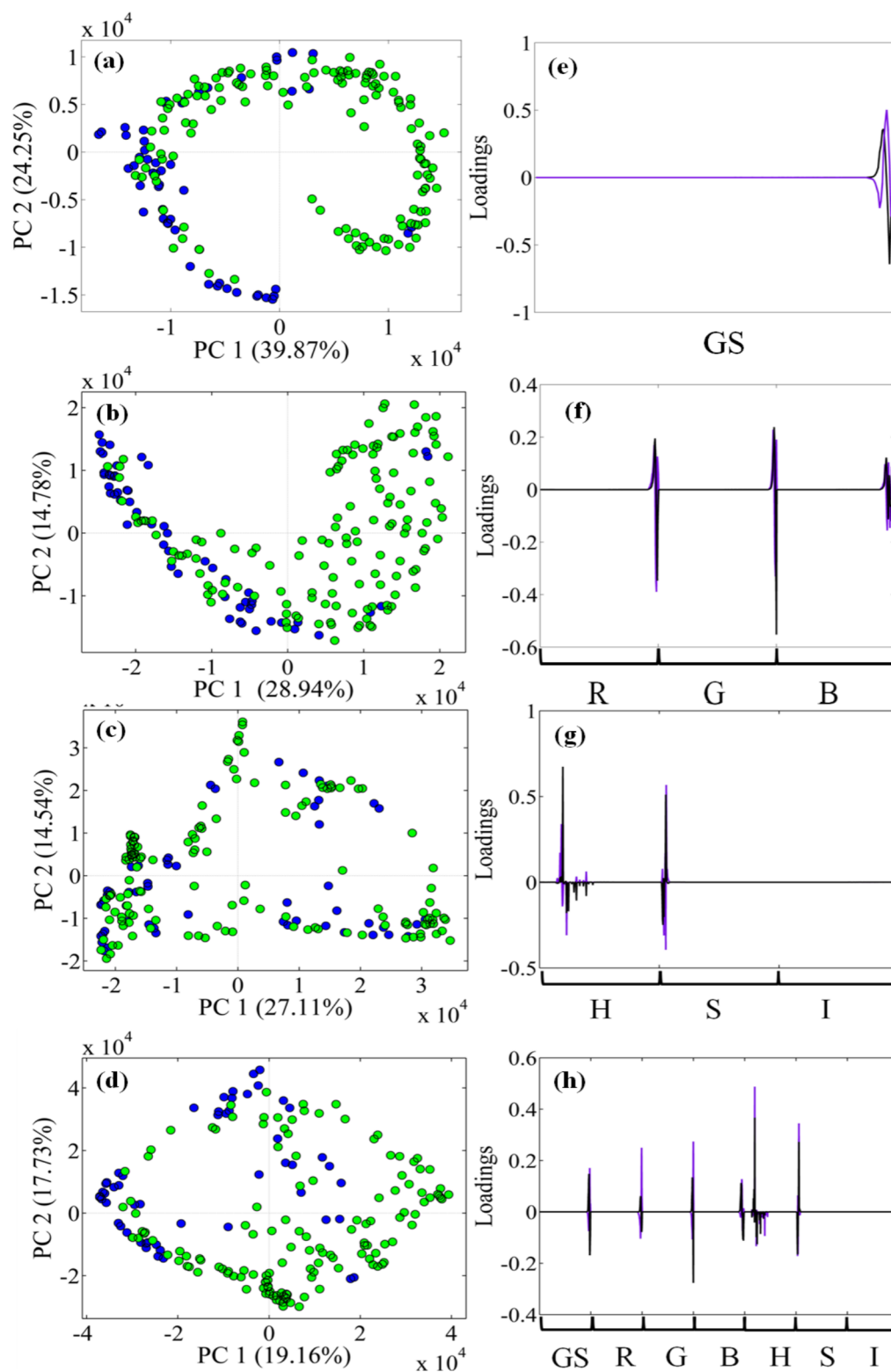
196 3.1. Análise exploratória dos dados

197

198 Para análise exploratória inicial dos dados, aplicou-se a Análise de Componentes
 199 Principais (PCA), cujos gráficos dos escores (*scores*) e pesos (*loadings*) de PC1 e PC2 estão
 200 ilustrados na **Figura 2** para os histogramas GS, RGB, HSI e GS+RGB+HSI.

201 Analisando as **Figuras 2a-d**, foi possível verificar que apenas para o histograma RGB
202 há uma tendência de agrupamento das amostras, embora ainda apresentando uma considerável
203 sobreposição entre as classes estudadas. É preciso considerar também que, exceto para o
204 histograma de cor GS, as duas PCs explicam menos de 50% da variância total dos dados e que
205 são necessárias mais que 10 PCs para explicar 90% da variância (**Figura S1 – Material**
206 **Suplementar**). Este resultado pode ser relacionado à alta variabilidade intrínseca das
207 matérias-primas em questão e da amostragem abrangente realizada neste trabalho, conforme
208 já detalhado na **Tabela 1**.

209 Neste contexto, vale ressaltar que os leites caprino e bovino apresentam-se como
210 líquidos opacos e turvos, enquanto suspensão coloidal, e cujas micelas de caseína e glóbulos
211 de gordura podem causar, a depender do tamanho e da densidade, diferentes níveis de
212 espalhamento da luz. Logo, isso impacta a opacidade das amostras e, consequentemente, na
213 variabilidade da distribuição de frequências ao longo dos diferentes canais de cor
214 (**HOLROYD, 2013; KUCHERYAVSKIY, MELENTEVA e BOGOMOLOV, 2014; SINGH e**
215 **GALLIER, 2017; CHEEMA et al., 2018**). É importante mencionar, ainda, que diferentes
216 fatores podem contribuir para a variabilidade do tamanho e da composição dessas partículas,
217 tais como, espécie, raça, estágio de lactação e características do manejo (por exemplo,
218 alimentação e quantidade de ordenhas diárias) (**SINGH e GALLIER, 2017; CHEEMA et al.,**
219 **2018**). Desta forma, considerando que algumas destas fontes de variabilidade intrínseca das
220 matérias-primas estão contempladas nas amostras utilizadas neste trabalho, a influência destas
221 sobre as informações analíticas obtidas por imagens digitais tornam a identificação de
222 adulteração por mistura de matrizes lácteas de elevada similaridade um desafio.



223
 224 **Figura 2.** Análise de Componentes Principais: gráficos (a, b, c, d) de escores dos histogramas de cor GS, RGB,
 225 HSI e GS+RGB+HSI, respectivamente e (e, f, g, h) pesos (loadings) obtidos para as duas primeiras componentes
 226 principais (PC1 – linha preta e PC2 – linha roxa). As amostras puras de leite caprino são representadas em azul e
 227 as adulteradas com leite bovino em verde.

228 Além disso, a observação dos gráficos de pesos das variáveis em PC1 e PC2 (**Figuras**
229 **2e-h**) para todos os histogramas e combinações testadas corroboram o fato de que a
230 informação analítica necessária à discriminação das amostras está concentrada nos valores dos
231 componentes de cor associados ao branco refletido das amostras. No canal RGB, por
232 exemplo, sabe-se que os tons de branco estão mais concentrados nas misturas equitativas de
233 valores de R, G e B próximos a 255. Contudo, como a possível diferenciação entre os dois
234 tipos de leite pode estar associada à presença de β -caroteno no leite bovino e à ausência desta
235 substância no leite caprino (que a converte totalmente em vitamina A) (CLAEYS et al., 2014;
236 BRANDÃO, ANJOS e BELL, 2017), nota-se que os pesos das variáveis em R e G são
237 maiores que em B. Isto acontece porque a cor percebida visualmente é a cor complementar da
238 luz absorvida. Em outras palavras, a coloração amarelada refletida acontece devido à absorção
239 de luz vermelha e verde. Portanto, tendo em vista que apenas PCA não foi capaz de separar as
240 classes de amostras puras e adulteradas, os classificadores “one-class” OC-PLS e DD-SIMCA
241 foram empregados e seus resultados são discutidos na seção seguinte.

242

243 3.2. Identificação de adulteração por classificadores “one-class”

244 A **Tabela 2** apresenta os resultados dos classificadores “one-class” para identificação de
245 adulteração de leite caprino em termos de sensibilidade, especificidade e eficiência usando os
246 diferentes histogramas e suas combinações.

247

Tabela 2. Resultados de sensibilidade, especificidade e eficiência para OC-PLS e DD-SIMCA nos conjuntos de treinamento e teste.

Histogramas de cor		OC-PLS				DD-SIMCA			
		LVs	Sens	Espec	Efic	PCs	Sens	Espec	Efic
GS	Treinamento	11	0,92	-	-	16	1,00	-	-
	Teste		0,93	0,65	0,78		0,71	0,75	0,73
RGB	Treinamento	17	1,00	-	-	17	1,00	-	-
	Teste		1,00	0,46	0,68		0,93	0,87	0,90
HSI	Treinamento	12	0,97	-	-	14	1,00	-	-
	Teste		0,93	0,45	0,65		0,93	0,79	0,86
GS+RGB	Treinamento	18	1,00	-	-	14	1,00	-	-
	Teste		1,00	0,45	0,67		0,93	0,77	0,85
GS+HSI	Treinamento	18	1,00	-	-	14	1,00	-	-
	Teste		1,00	0,54	0,74		1,00	0,76	0,87
RGB+HSI	Treinamento	14	0,97	-	-	16	1,00	-	-
	Teste		0,93	0,51	0,69		0,86	0,86	0,82
GS+RGB+HSI	Treinamento	12	0,97	-	-	12	1,00	-	-
	Teste		0,93	0,43	0,64		0,93	0,79	0,86

Nota: Sens – sensibilidade; Espec – especificidade; Efic – eficiência; PCs – componentes principais; LVs – variáveis latentes.

Preliminarmente, foi possível observar que para alcançar os valores mais elevados de sensibilidade do modelo, os algoritmos OC-PLS e DD-SIMCA selecionaram acima de 10 variáveis latentes e componentes principais, respectivamente, em todos os histogramas de cor avaliados. Tais resultados estão em concordância com a análise exploratória realizada por PCA e também podem ser associados à complexidade e à variabilidade intrínseca da matriz estudada (leite caprino), mesmo quando apenas a classe alvo é considerada na construção do modelo.

Ainda de acordo com os resultados apresentados na **Tabela 2**, para OC-PLS verificou-se que, apesar dos resultados de sensibilidade serem elevados (acima de 0,92), para o conjunto de treinamento em todos os histogramas de cor, apenas para canal analítico GS foi observada

263 eficiência superior a 0,75 no conjunto de teste. Tais valores são derivados das baixas taxas de
264 especificidade (menores que 0,65), relativa ao elevado número de erros por inclusão de
265 amostras adulteradas na classe alvo. Quanto ao histograma de cor selecionado, uma vez que
266 GS pode ser resultante da medida de intensidade de luz em cada pixel, a sua contribuição na
267 diferenciação entre os leites caprino e bovino pode ser devida à ausência de pigmentos
268 amarelos no leite caprino e presença destes no leite bovino. Em outras palavras, a sua
269 contribuição pode estar relacionada à aparência mais branca do leite caprino.

270 Para o algoritmo DD-SIMCA, o melhor desempenho em termos de eficiência foi
271 obtido para o histograma RGB, que alcançou 0,90 no conjunto de teste, evidenciando o
272 reduzido número de erros nas classes alvo (amostras puras) e não-alvo (amostras adulteradas).
273 Neste caso, a informação contida na região RGB está correlacionada à percepção da cor
274 amarelada nas amostras adulteradas, tendo em vista que a cor do leite bovino pode variar de
275 acordo com o teor de β -caroteno (CLAEYS et al., 2014; ALICHANIDIS; MOATSOU e
276 POLYCHRONIADOU, 2016; BRANDÃO, ANJOS e BELL, 2017) e, por este motivo, pode
277 fornecer informações para diferenciação entre as classes.

278 A **Figura 3** apresenta os gráficos de área de aceitação para OC-PLS e DD-SIMCA nos
279 conjuntos de treinamento e teste para os histogramas de cor nos quais os melhores
280 desempenhos foram observados (GS e RGB, respectivamente).

281

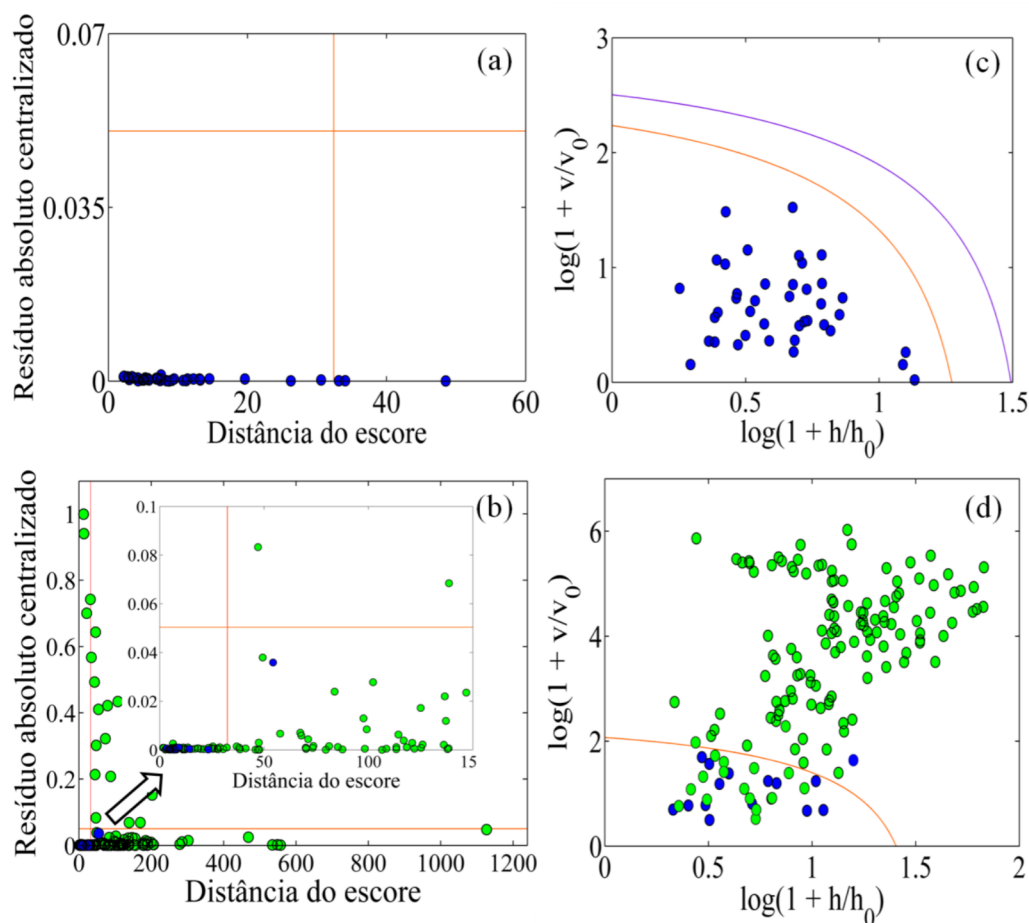


Figura 3. Gráficos de área de aceitação para o conjunto de (a,c) treinamento e (b,d) teste empregando OC-PLS para o histograma GS e DD-SIMCA para histograma RGB, respectivamente. As bolas azuis e verdes representam amostras de leite caprino puras e adulteradas, respectivamente. As linhas laranjas representam a borda entre as amostras regulares e extremas; a linha púrpura é a fronteira da área de aceitação.

Para OC-PLS, as amostras são descritas em termos do resíduo absoluto centralizado e da distância do escore. Os limiares são definidos em termos de cada parâmetro individualmente, definindo uma área de aceitação retangular (linhas laranjas nas Figuras 3a e 3b) (XU et al., 2013; ARAÚJO, MARINHO e GOMES, 2018). Na Figura 3a é possível observar que três amostras estão fora dos limites estabelecidos para o conjunto de treinamento. Já no conjunto de teste foram observados 49 erros, nos quais uma amostra da classe alvo (leite caprino puro) foi identificada como não pertencente ao grupo e 48 amostras adulteradas foram atribuídas à classe alvo.

Já DD-SIMCA utiliza o conceito de amostras extremas na definição da área de aceitação robusta para o conjunto de teste. As amostras são descritas em termos da variável c , a qual combina distâncias ortogonais e de escores das amostras. Assume-se, então, que a variável c possui distribuição qui-quadrado e dois limiares são estabelecidos. Amostras com valores de c menores que o primeiro limiar (linha laranja nas Figuras 3c e 3d) são típicas da classe alvo, enquanto amostras com valores de c maiores que o primeiro limiar e menores que o segundo pertencem à classe alvo, mas exibem comportamento extremo. Amostras com valores de c maiores que o segundo limiar (linha púrpura na Figura 3c) são *outliers* (ZONTOV et al., 2017; ARAÚJO; MARINHO e GOMES, 2018). Para o conjunto de treinamento, não foram detectadas amostras extremas e *outliers* empregando DD-SIMCA, independente do histograma de cor avaliado. Contudo, para o conjunto de teste, o melhor resultado foi obtido para o histograma RGB que detectou 19 erros de classificação, em que uma amostra leite caprino puro foi identificada como adulterada e 18 amostras adulteradas foram atribuídas à classe alvo. Esses resultados apresentam performance superior aos encontrados por Santos e Pereira-Filho (2013), que empregaram imagens digitais e SIMCA para identificação de adulteração em leite bovino por adição individual e combinada de 5 adulterantes (água, soro de leite, peróxido de hidrogênio e urina e leite sintéticos), obtendo valores de sensibilidade de 85 e 80% para as classes de amostras puras e adulteradas, respectivamente, no conjunto de teste.

Portanto, considerando-se a performance geral do modelo DD-SIMCA usando o histograma RGB, das 54 amostras de leite caprino puro, apenas uma foi classificada equivocadamente, o que representa uma taxa de classificação correta de 98% das amostras estudadas. Por outro lado, apesar de apenas 120 das 138 amostras adulteradas serem classificadas corretamente (87% de acerto), a metodologia desenvolvida representa uma

320 alternativa importante para a finalidade proposta, dada a complexidade e variabilidade
321 intrínseca dos leites caprino e bovino aqui estudados.

322 Não obstante, uma vez que o sistema de produção de leite caprino no Brasil é
323 constituído por pequenos produtores, a construção de modelos individuais poderia resultar no
324 aumento do desempenho de classificação, tendo em vista a redução da variabilidade. Sendo
325 assim, tal metodologia poderia funcionar como uma ferramenta de *screening in situ* na
326 recepção dos leites de pequenos produtores pelas cooperativas.

327

328 **4. CONCLUSÕES**

329 Neste trabalho nós demonstramos a viabilidade da utilização de imagens digitais como
330 ferramenta para identificação de adulteração de leite caprino por adição de leite bovino.
331 Considerando a complexidade da matéria-prima, as características deste tipo de adulteração e
332 a acessibilidade da metodologia apresentada, dado seu baixo custo e facilidade de operação,
333 tal ferramenta analítica apresenta potencial para aplicabilidade na triagem da autenticidade de
334 amostras de leite caprino por pequenos produtores e cooperativas.

335

336 **AGRADECIMENTOS**

337 Os autores agradecem ao NUPEA-UEPB pelo suporte nas análises de referência e aos
338 produtores dos leites caprinos e bovinos pelo fornecimento das amostras. David Fernandes e
339 Mário Araújo agradecem ao CNPq pelas bolsas de pesquisa.

340

341 **REFERÊNCIAS**

342 AGRIMONTI, C.; PIRONDINI, A.; MARMIROLI, M.; MARMIROLI, N. (2015) A
343 quadruplex PCR (qPCR) assay for adulteration in dairy products. Food Chemistry, 187, 58–
344 64. <https://doi.org/10.1016/j.foodchem.2015.04.017>

- 345 ALICHANIDIS, E.; MOATSOU, G.; POLYCHRONIADOU, A. (2016) Composition and
346 Properties of Non-cow Milk and Products. In *Non-Bovine Milk and Milk Products*. Elsevier
347 Inc. <https://doi.org/10.1016/B978-0-12-803361-6.00005-3>
- 348 AHMED, A. S.; EL-BASSIONY, T.; ELMALT, L. M.; IBRAHIM, H. R. (2015)
349 Identification of potent antioxidant bioactive peptides from goat milk proteins. **Food**
350 **Research International**, 74, 80–88. <https://doi.org/10.1016/j.foodres.2015.04.032>
- 351 AOAC (2005). Official Methods of Analysis of Association of Official Analytical Chemists.
352 18th Edition, Washington, DC.
- 353 ARAÚJO, A.; MARINHO, W.; GOMES, A. A. (2018) A fast and inexpensive chemometric-
354 assisted method to identify adulteration in açai (*Euterpe oleracea*) using digital images. **Food**
355 **Analytical Methods**, 11, 1920-1926. <https://doi.org/10.1007/s12161-017-1127-4>
- 356 AZEEM, S. M. A.; MADBOULY, M. D.; EL-SHARAT, M. F. (2019) Determination of
357 nitrite in processed meat using digital image method and powdered reagent. **Food**
358 **Composition Analysis**, v. 81, p. 28-36. <https://doi.org/10.1016/j.jfca.2019.05.003>
- 359 BOTELHO, B. G.; ASSIS, L. P.; SENA, M. M. (2014) Development and analytical
360 validation of a simple multivariate calibration method using digital scanner images for sunset
361 yellow determination in soft beverages. **Food Chemistry**, 159, 175-180.
362 <http://dx.doi.org/10.1016/j.foodchem.2014.03.048>
- 363 BOTELHO, B. G.; DANTAS, K. C. F.; SENA, M. M. (2017) Determination of allura red dye
364 in hard candies by using digital images obtained with a mobile phone and N-PLS.
365 **Chemometrics and Intelligent Laboratory Systems**, 167, 44-49.
366 <http://dx.10.1016/j.chemolab.2017.05.004>

- 367 BRANDÃO, M. P.; ANJOS, V. C.; BELL, M. J. V. (2017) Time resolved fluorescence of
368 cow and goat milk powder. **Spectrochimica Acta Part A: Molecular and Biomolecular**
369 **Spectroscopy**, 171, 193–199. <https://doi.org/10.1016/j.saa.2016.08.007>.
- 370 CHEEMA, M.; SMITH, P. B.; PATTERSON, A. D.; HRISTOV, A.; MARTE, F. M. (2018)
371 The association of lipophilic phospholipids with native bovine casein micelles in skim milk:
372 Effect of lactation stage and casein micelle size. **Journal of Dairy Science**, 101,1-16.
373 <https://doi.org/10.3168/jds.2017-14137>
- 374 CHOOPAN, R.; THANAKIATKRAI, P.; KITPIPIT, T. (2017) Simultaneous species
375 identification in milk and dairy products using direct pcr. **Forensic Science International:**
376 **Genetics Supplement Series**, 6, 214-215. <http://dx.doi.org/10.1016/j.fsigs.2017.09.077>
- 377 CLAEYS, W. L.; VERRAES, C.; CARDOEN, S.; DE BLOCK, J.; HUYGHEBAERT, A.;
378 RAES, K.; DEWETTINCK, K.; HERMAN, L. (2014) Consumption of raw or heated milk
379 from different species: An evaluation of the nutritional and potential health benefits. **Food**
380 **Control**, 42, 188-201. <http://dx.doi.org/10.1016/j.foodcont.2014.01.045>
- 381 CLARK, S.; GARCÍA, M. B. M. (2017) A 100-Year Review: Advances in goat milk
382 research. **Journal of Dairy Science**, 100(12), 10026–10044. [https://doi.org/10.3168/jds.2017-](https://doi.org/10.3168/jds.2017-13287)
383 [13287](https://doi.org/10.3168/jds.2017-13287)
- 384 COSTA, G. B.; FERNANDES, D. D. S.; ALMEIDA, V. E.; ARAÚJO, T. S. P.; MELO, J.
385 P.; DINIZ, P. H. G. D.; VÉRAS, G. (2015) Digital image-based classification of biodiesel.
386 **Talanta**, 139, 50-55. <http://dx.doi.org/10.1016/j.talanta.2015.02.043>
- 387 DI DOMENICO, M.; DI GIUSEPPE, M.; RODRÍGUEZ, J. D. W.; CAMMÀ, C. (2017)
388 Validations of a fast real-time PCR method to detect fraud and mislabeling in milk and dairy
389 products. **Journal of Dairy Science**, 100, 106-112. <https://doi.org/10.3168/jds.2016-11695>

- DI PINTO, A.; TERIO, V.; MARCHETTI, P.; BOTTARO, M.; MOTTOLA, A.; BOZZO, G.;
BONERBA, E.; CECI, E.; TANTILLO, G. (2017) DNA-based approach for species
identification of goat-milk products. *Food Chemistry*, 229, 93–97.
<https://doi.org/10.1016/j.foodchem.2017.02.067>
- DINIZ, P. H. G. D.; DANTAS, H. V.; MELO, K. D. T.; BARBOSA, M. F.; HARDING, D.
P.; NASCIMENTO, E. C. L.; PISTONESI, M. F.; BAND, B. S. F.; ARAÚJO, M. C. U.
(2012) Using a simple digital câmera and SPA-LDA modeling to screen teas. **Analytical
Methods**, 4, 2648-2652. <http://dx.doi.org/10.1039/C2AY25481F>
- EL-SAYED, S.; GALAL, H.; ALLAN, A.; SCHMIDHALTER, U. (2016) Passive reflectance
sensing and digital image analysis for assessing quality parameters of mango fruits. **Scientia
Horticulturae**, 212, 136-147. <http://dx.doi.org/10.1016/j.scienta.2016.09.046>
- FERNANDES, D. D. S.; ROMEO, F.; KREPPER, G.; DI NEZIO, M. S.; PISTONESI, M. F.;
CENTURIÓN, M. E.; ARAÚJO, M. C. U.; DINIZ, P. H. G. D. (2019) Quantification and
identification of adulteration in the fat content of chicken hamburgers using digital images
and chemometric tools. *LWT - Food Science and Technology*, 100, 20–27.
<https://doi.org/10.1016/j.lwt.2018.10.034>.
- GOLINELLI, L. P.; CARVALHO, A. C.; CASAES, R. S.; LOPES, C. S. C.; DELIZA, R.;
PASCHOALIN, V. M. F.; SILVA, J. T. (2014) Sensory analysis and species-specific PCR
detect bovine milk adulteration of frescal (fresh) goat cheese. **Journal of Dairy Science**,
97(11), 6693–6699. <https://doi.org/10.3168/jds.2014-7990>
- HANDFORD, C. E.; CAMPBELL, K.; ELLIOTT, C. T. (2016) Impacts of milk fraud on food
safety and nutrition with special emphasis on developing countries. **Comprehensive Reviews
in Food Science and Food Safety**, 15, 130-142. <http://dx.doi:10.1111/1541-4337.12181>

- 413 HOLROYD, S. E. (2013) The use of near infrared spectroscopy on milk and milk products.
414 **Journal of Near Infrared Spectroscopy**, 21, 311-322. <https://doi.org/10.1255/jnirs.1055>
- 415 KENDALL, H.; CLARK, B.; RHYMER, C.; KUZNESOF, S.; HAJLSLOVA, J.;
416 TOMANIOVA, M.; BRERETON, P.; FREWER, L. (2019) A systematic review of consumer
417 perceptions of food fraud authenticity: A European perspective. **Trends in Food Science &**
418 **Technology**, 94, 79-90. <https://doi.org/10.1016/j.tifs.2019.10.005>
- 419 KIM, M. J.; KIM, H. Y. (2018) Direct duplex real-time loop mediated isothermal
420 amplification assay for the simultaneous of cow and goat species origin of milk and yogurt
421 products for field use. **Food Chemistry**, 246, 26-31.
422 <http://dx.doi.org/10.1016/j.foodchem.2017.11.014>
- 423 KUCHERYAVSKIY, S.; MELENTEVA, A.; BOGOMOLOV, A. (2014) Determination of
424 fat and total protein content in milk using conventional digital imaging. **Talanta**, 121, 144-
425 152. <http://dx.doi.org/10.1016/j.talanta.2013.12.055>
- 426 LAVINE, B. K. (2009). Validation of Classifiers. **Comprehensive Chemometrics**, 3, 587-
427 599. <https://doi.org/10.1016/B978-044452701-1.00027-2>
- 428 LIMA, C. M.; FERNANDES, D. D. S.; PEREIRA, G. E.; GOMES, A. A.; ARAÚJO, M. C.
429 U.; DINIZ, P. H. G. D. (2020) Digital image-based tracing of geographic origin, winemaker,
430 and grape type for red wine authentication. **Food Chemistry**, 312, 126060.
431 <https://doi.org/10.1016/j.foodchem.2019.126060>
- 432 MASSART, D. L.; VANDEGINSTE, B. G. M.; BUYDENS, L. M. C.; DE JONG, S.; LEWI,
433 P. J.; SMEYERS-VERBEKE, J. (1998) Handbook of Chemometrics and Qualimetrics, Part
434 A. Elsevier.

- 435 MCANDREW, A. (2015) A computational introduction to digital image processing. 2^a Ed,
436 CRC Press.
- 437 MOORE, J. C.; SPINK, J.; LIPP, M. (2012) Development and application of a database of
438 food ingredient fraud and economically motivated adulteration from 1980 to 2010. **Journal of**
439 **Food Science**, 77, 4, 118-126. <http://dx.doi:10.1111/j.17503841.2012.02657.x>
- 440 NASCIMENTO, C. F.; SANTOS, P. M.; PEREIRA-FILHO, E. R.; ROCHA, F. R. P. (2017)
441 Recent advances on determination of milk adulterants. **Food Chemistry**, 221, 1232–1244.
442 <https://doi.org/10.1016/j.foodchem.2016.11.034>
- 443 OLIVERI, P.; DOWNEY, G. (2012) Multivariate class modeling for the verification of food-
444 authenticity claims. **TrAC - Trends in Analytical Chemistry**, 35, 74–86.
445 <https://doi.org/10.1016/j.trac.2012.02.005>
- 446 PESIC, M.; BARAC, M.; VRVIC, M.; RISTIC, N.; MACEJ, O.; STANOJEVIC, S. (2011)
447 Qualitative and quantitative analysis of bovine milk adulteration in caprine and ovine milks
448 using native-PAGE. **Food Chemistry**, 125(4), 1443–1449.
449 <https://doi.org/10.1016/j.foodchem.2010.10.045>
- 450 RODRIGUES, N. P. A.; GIVISIEZ, P. E. N.; QUEIROGA, R. C. R. E.; AZEVEDO, P. S.;
451 GEBREYES, W. A.; OLIVEIRA, C. J. B. (2012) Milk adulteration: Detection of bovine milk
452 in bulk goat milk produced by smallholders in northeastern Brazil by a duplex PCR assay.
453 **Journal of Dairy Science**, v. 95 (5), 2749-2752.
- 454 SINGH, H.; GALLIER, S. (2017) Nature's complex emulsion: The fat globules of milk. **Food**
455 **Hydrocolloids**, 68, 81-89. <http://dx.doi.org/10.1016/j.foodhyd.2016.10.011>

- 456 SANTOS, P. M. S.; PEREIRA-FILHO, E. R. (2013) Digital image analysis- na alternative
457 tool for monitoring milk authenticity. **Analytical Methods**, 5, 3669-3674.
458 <http://dx.doi.org/10.1039/c3ay40561c>
- 459 SONG, H.; XUE, H.; HAN, Y. (2011) Detection of cow's milk in Shaanxi goat's milk with an
460 ELISA assay. **Food Control**, 22(6), 883–887. <https://doi.org/10.1016/j.foodcont.2010.11.019>
- 461 SOUTO, U. T. C. P.; BARBOSA, M. F.; DANTAS, H. V.; PONTES, A. S.; LYRA, W. S.;
462 DINIZ, P. H. G. D.; ARAÚJO, M. C. U; SILVA, E. C. (2014) Screening for coffee
463 adulteration using digital images and SPA-LDA. **Food Analytical Methods**, 8, 1515-1521.
464 <http://dx.10.1007/s12161-014-0020-7>
- 465 VERRUCK, S., DANTAS, A., PRUDENCIO, E. S. (2019) Functionality of the components
466 from goat's milk, recent advances for functional dairy products development and its
467 implications on human health. **Journal of Functional Foods**, 52, 243–257.
468 <https://doi.org/10.1016/j.jff.2018.11.017>
- 469 VIEITEZ, I.; IRIGARAY, B.; CALLEJAS, N.; GONZÁLEZ, V.; GIMENEZ, S.;
470 ARECHAVALETA, A.; GROMPONE, M.; GÁMBARO, A. (2016) Composition of fatty
471 acids and triglycerides in goat cheeses and study of the triglyceride composition of goat milk
472 and cow milk blends. **Journal of Food Composition and Analysis**, 48, 95–101.
473 <https://doi.org/10.1016/j.jfca.2016.02.010>
- 474 XU, L.; YAN, S. M.; CAI, C. B.; YU, X. P. (2013) One-Class partial least squares (OCPLS)
475 classifier. **Chemometrics and Intelligente Laboratory Systems**, 126, 1-5.
476 <http://dx.doi.org/10.1016/j.chemolab.2013.04.008>

- 477 ZHANG, W.; XUE, J. (2016) Economically motivated food fraud and adulteration in China:
478 An analysis based on 1553 media reports. **Food Control**, 67, 192-198.
479 <http://dx.doi.org/10.1016/j.foodcont.2016.03.004>
- 480 ZONTOV, Y. V.; RODIONOVA, O. Y.; KUCHERYAVSKIY, S. V.; POMERANTSEV, A.
481 L. (2017) DD-SIMCA – A MATLAB GUI tool for data driven SIMCA approach.
482 **Chemometrics and Intelligente Laboratory Systems**, 167, 23-28.
483 <http://dx.doi.org/10.1016/j.chemolab.2017.05.010>
- 484

Apêndice A. Material Suplementar

Imagens digitais como ferramenta para autenticação de leite caprino em termos de adulteração por adição de leite bovino: um estudo de viabilidade

Elainy Virginia dos Santos Pereira^{a,b}, David Douglas de Sousa Fernandes^c,
Mário César Ugulino de Araújo^c, Paulo Henrique Gonçalves Dias Diniz^{d*}, Maria Inês
Sucupira Maciel^{e*}

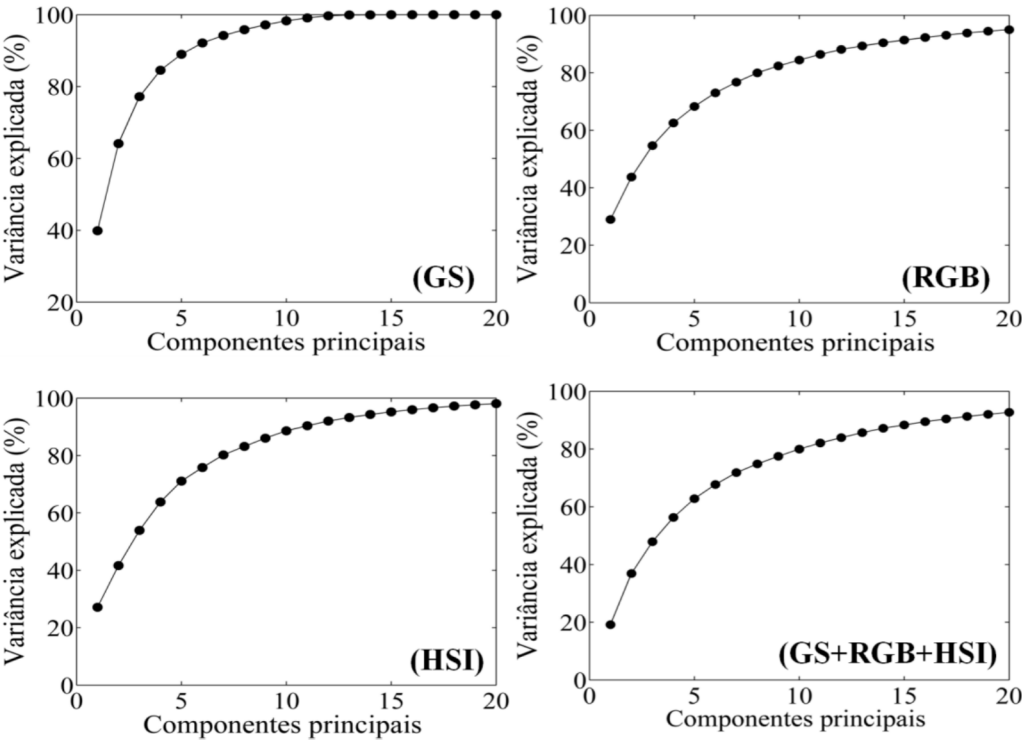
^a Programa de Pós-graduação em Ciência e Tecnologia de Alimentos – PPGCTA Centro de Tecnologia, Universidade Federal da Paraíba, CEP- 58051-900, João Pessoa, Paraíba, Brasil

^b Núcleo de Pesquisa e Extensão em Alimentos - NUPEA, Centro de Ciências e Tecnologia, Universidade Estadual da Paraíba, CEP 58429-500, Campina Grande, Paraíba, Brasil

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

^d Programa de Pós-graduação em Química Pura e Aplicada, Universidade Federal do Oeste da Bahia, CEP 47810-059, Barreiras, Bahia, Brasil

^e Programa de Pós-graduação em Ciência e Tecnologia de Alimentos, Universidade Federal Rural de Pernambuco, CEP 52171-900, Recife, Pernambuco, Brasil



507

508 **Figura S1** – Gráficos das variâncias explicadas em função das componentes principais, para os histogramas de
509 cor GS (escala de cinza), RGB (vermelho, verde e azul), HSI (matiz, saturação e intensidade) individuais e
510 combinados (GS+RGB+HSI).

5 CONSIDERAÇÕES FINAIS

Diante dos resultados obtidos nos artigos apresentados conclui-se que:

- Metodologias analíticas baseadas na espectroscopia NIR combinadas com ferramentas quimiométricas são adequadas para identificação de adulteração de leite caprino por adição de leite bovino, mesmo quando baixos níveis de adulteração (de até 1% m m⁻¹) são investigados. Além disso, a depender do equipamento utilizado e da resolução dos dados, as informações obtidas podem ser modeladas para quantificar simultaneamente o percentual de adulteração e os teores de proteínas e lipídeos totais, como é o caso do espectrofotômetro NIR convencional de bancada.

- A adequabilidade de um espectrofotômetro NIR miniaturizado portátil para identificação de adulteração de leite caprino torna esta metodologia uma ferramenta viável para realização de *screening in situ* da autenticidade de leite caprino.

- A metodologia baseada em Imagens Digitais, apesar de ter alcançado uma eficiência inferior às metodologias baseadas em NIRS, pode ser considerada uma ferramenta analítica promissora com potencial de aplicabilidade na triagem da autenticidade de amostras de leite caprino por pequenos produtores e cooperativas, principalmente se modelos locais/individuais forem construídos, uma vez que as fontes de variabilidade seriam reduzidas.

Ante o exposto, considerando a complexidade e a variabilidade intrínseca da matéria-prima estudada e que a adulteração de leite caprino por adição de leite bovino é caracterizada pela mistura de matrizes alimentícias, de tal forma que sua identificação representa um desafio analítico, as metodologias desenvolvidas neste trabalho demonstraram-se adequadas para a finalidade proposta.

Além disso, é necessário considerar que, para além das vantagens como redução de custos e tempo de análise, a importância principal das metodologias apresentadas relaciona-se com o desenvolvimento de métodos analíticos verdes aplicáveis à tecnologia de alimentos, com perspectiva à extensão da empregabilidade para derivados de leite caprino, bem como para outras matrizes alimentícias.