



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I - CAMPINA GRANDE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA
CURSO DE MESTRADO EM QUÍMICA**

HELLEN JAINNE DO NASCIMENTO PEREIRA

**DESENVOLVIMENTO DE METODOLOGIA PORTÁTIL E NÃO DESTRUTIVA
PARA DETERMINAÇÃO DE ADULTERAÇÃO EM LEITE DE CABRA**

**CAMPINA GRANDE - PB
2025**

HELLEN JAINNE DO NASCIMENTO PEREIRA

**DESENVOLVIMENTO DE METODOLOGIA PORTÁTIL E NÃO DESTRUTIVA
PARA DETERMINAÇÃO DE ADULTERAÇÃO EM LEITE DE CABRA**

Dissertação apresentada à Coordenação do Curso de
Programa de Pós-Graduação em Química - Mestrado
da Universidade Estadual da Paraíba, como requisito
parcial à obtenção do título de Mestra em Química.

Área de Concentração: Química

Linha de Pesquisa: Metodologias Analíticas
Avançadas

Orientador: Prof. Dr. Paulo Henrique Dias Gonçalves Diniz

Coorientador: Prof. Dr. David Douglas de Sousa Fernandes

**CAMPINA GRANDE - PB
2025**

É expressamente proibida a comercialização deste documento, tanto em versão impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que, na reprodução, figure a identificação do autor, título, instituição e ano do trabalho.

P436d Pereira, Hellen Jainne do Nascimento.

Desenvolvimento de metodologia portátil e não destrutiva para determinação de adulteração em leite de cabra [manuscrito] / Hellen Jainne do Nascimento Pereira. - 2025.

64 f. : il. color.

Digitado.

Dissertação (Programa de Pós-Graduação em Química - Mestrado) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2025.

"Orientação : Prof. Dr. Paulo Henrique Gonçalves Dias Diniz, Professor".

1. Derivados lácteos. 2. Controle de qualidade do leite. 3. Análise de alimentos. 4. Espectroscopia vibracional. 5. Química analítica verde. I. Título

21. ed. CDD 543

HELLEN JAINNE DO NASCIMENTO PEREIRA

DESENVOLVIMENTO DE METODOLOGIA PORTÁTIL E NÃO DESTRUTIVA
PARA DETERMINAÇÃO DE ADULTERAÇÃO EM LEITE DE CABRA

Dissertação apresentada à
Coordenação do Curso de Programa de
Pós-Graduação em Química - Mestrado
da Universidade Estadual da Paraíba,
como requisito parcial à obtenção do
título de Mestra em Química

Linha de Pesquisa: Metodologias
Analíticas Avançadas - Campina
Grande.

Aprovada em: 18/07/2025.

BANCA EXAMINADORA

Documento assinado eletronicamente por:

- **Valeria Haydee Springer** (***.088.382-**), em **18/08/2025 10:46:13** com chave **b4bb5f247c3911f090171a7cc27eb1f9**.
- **Rennan Pereira de Gusmão** (***.267.514-**), em **18/08/2025 16:13:31** com chave **6db526187c6711f0b57906adb0a3afce**.
- **Paulo Henrique Gonçalves Dias Diniz** (***.617.384-**), em **18/08/2025 10:36:50** com chave **6503f7a87c3811f0b6291a7cc27eb1f9**.

Documento emitido pelo SUAP. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse https://suap.uepb.edu.br/comum/autenticar_documento/ e informe os dados a seguir.

Tipo de Documento: Folha de Aprovação do Projeto Final

Data da Emissão: 18/08/2025

Código de Autenticação: 16f793



Aos meus avós, Dona Zefa, Dona Zefinha e Seu Luiz,todos em memória, que enfrentaram com coragem e dignidade as durezas do Sertão. Que esta conquista também lhes pertença.

AGRADECIMENTOS

A Deus, pela força, saúde e sabedoria concedidas ao longo desta jornada, guiando-me em cada decisão e iluminando meu caminho.

À minha família, especialmente minha mãe Lucimar, meu pai Elenildo, minha irmã Hemanuelle, meu irmão Rafael, meu avô José Pereira e minha tia Maria José, agradeço pelo amor, apoio e incentivo, que foram minha base e inspiração durante toda essa caminhada.

Ao meu orientador, Prof. Dr. Paulo Henrique Dias Gonçalves Diniz, pela paciência, orientação e dedicação ao longo deste trabalho. Sua expertise, comprometimento, acolhimento e carinho tornaram a jornada mais leve e enriquecedora para meu desenvolvimento acadêmico e pessoal.

Ao meu coorientador, Dr. David Douglas de Sousa Fernandes, pela colaboração, valiosas contribuições e apoio sempre disponível, que enriqueceram imensamente este trabalho. As conversas e risadas sinceras tornaram tudo muito mais agradável.

Agradeço à banca de defesa, Prof^a. Dr^a. Valeria Haydée Springer e Prof. Dr. Rennan Pereira de Gusmão, pelos comentários construtivos, questionamentos instigantes e pelo comprometimento em avaliar este trabalho com rigor científico e generosidade. As sugestões feitas enriqueceram significativamente a qualidade da pesquisa e contribuirão para meu crescimento acadêmico e profissional.

À banca de qualificação, Dr^a. Welma Thaíse Silva Vilar e Dr^a. Emanuella Santos Sousa, pela generosidade em participar de uma etapa tão importante e pelas contribuições construtivas que aprimoraram este estudo.

Ao Diego Leam, pelo amor, paciência e apoio incondicional em todos os momentos desta jornada. Sua presença foi fundamental para minha motivação e equilíbrio.

Aos amigos, que estiveram presentes nos momentos de alegria e nos desafios, oferecendo palavras de encorajamento e um companheirismo inestimável. Principalmente àqueles que a pós-graduação me presenteou. Em especial, agradeço a Luana Leal, companheira

de labuta e amiga inseparável, e a Aline Peres, que já considero a família que meu coração escolheu, pela amizade e apoio constantes.

A Tina Lima, que mesmo distante me ajudou imensamente, a Joseane Oliveira, amiga de sempre e para sempre. Obrigada minhas amigas!

Ao Programa de Pós-Graduação em Química, por proporcionar a estrutura necessária e o ambiente de aprendizado que tornaram possível o desenvolvimento deste trabalho.

Aos meus professores, deixo meu sincero agradecimento pela paciência, incentivo e valiosas contribuições que enriqueceram meu percurso no mestrado. A orientação e os ensinamentos recebidos foram essenciais para a realização deste trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

A todos que, direta ou indiretamente, contribuíram para a concretização deste sonho, o meu mais sincero agradecimento.

"Continue a nadar."

— Dory, Procurando Nemo

RESUMO

Este estudo apresenta uma abordagem inovadora para a determinação do teor de gordura e para a detecção de adulteração com leite de vaca em amostras de leite de cabra. Para isso, foi utilizado um espectrômetro de infravermelho próximo (NIR) miniaturizado, aliado a técnicas avançadas de calibração multivariada. Os modelos estatísticos aplicados baseiam-se no Algoritmo de Projeções Sucessivas (SPA) para seleção de variáveis e intervalos, sendo empregados em dois métodos: a Regressão Linear Múltipla (SPA-MLR) e os Mínimos Quadrados Parciais (iSPA-PLS). Entre as diferentes técnicas de pré-processamento testadas, a suavização de Savitzky-Golay com janela de 11 pontos (SGS) favoreceu o melhor desempenho preditivo dos modelos quantitativos. O modelo SGS/iSPA-PLS alcançou excelentes resultados: coeficientes de correlação (r_{pred}) de 0,97 e 0,99, erros quadráticos médios de predição (RMSEP) de 0,12 g/100 g e 2,15 g/100 g, razões de desempenho para desvio (RPD) de 4,32 e 8,96, e erros relativos de predição (REP) de 2,70% e 8,04%, respectivamente para a estimativa do teor de gordura e para a detecção da adulteração com leite de vaca. Essa metodologia se mostra eficaz para lidar com os principais desafios relacionados à variabilidade composicional do leite e à presença de adulterantes, oferecendo uma ferramenta analítica robusta, portátil e eficiente. Trata-se de uma solução que pode ser aplicada em pesquisas científicas e também no controle de qualidade do leite de cabra.

Palavras-Chave: derivados lácteos; controle de qualidade do leite; análise de alimentos; espectroscopia vibracional; química analítica verde.

ABSTRACT

This study proposes an innovative approach for determining fat content and detecting cow milk adulteration in goat milk samples using a miniaturized near-infrared (NIR) spectrometer combined with advanced multivariate calibration techniques. Statistical models based on the Successive Projections Algorithm (SPA) for variable and interval selection were applied through Multiple Linear Regression (SPA-MLR) and interval Partial Least Squares (iSPA-PLS) methods. Among various preprocessing techniques tested, the 11-point Savitzky-Golay smoothing (SGS) provided the best predictive performance for the quantitative models. The SGS/iSPA-PLS model achieved high correlation coefficients (r_{pred}) of 0.97 and 0.99, root mean square errors of prediction (RMSEP) of 0.12 g/100 g and 2.15 g/100 g, ratios of performance to deviation (RPD) of 4.32 and 8.96, and relative errors of prediction (REP) of 2.70% and 8.04%, respectively for fat content estimation and adulteration detection. This methodology effectively addresses challenges related to compositional variability and adulterants presence, providing a robust, portable, and efficient analytical tool suitable for scientific research and quality control of goat milk.

Keywords: Dairy products; Milk quality control; Food analysis; vibrational spectroscopy; green analytical chemistry.

LISTA DE ILUSTRAÇÕES

	Página
Figura 1 – Esquema do procedimento analítico para aquisição e tratamento de espectros NIR de amostras de leite.....	39
Figura 2 – Espectros médios de NIR pré-processados com OFF (Fig. 1a), LBC (Fig. 1b), MSC (Fig. 1c), SGS (Fig. 1d), SNV (Fig. 1e) e SGD (Fig. 1f) para as amostras de leite de cabra puro (linha sólida laranja) e adulterado (linha sólida azul). As áreas sombreadas ao redor dos espectros médios de NIR representam a variabilidade máxima observada em cada canal analítico para as duas classes estudadas	43
Figura 3 – Gráficos e resultados relacionados à determinação do teor de gordura em amostras de leite de cabra e suas misturas com leite de vaca, utilizando os espectros NIR pré-processados com suavização Savitzky-Golay e janela de 11 pontos (SGS): (a-c) Gráficos de influência, que ilustram as distâncias entre os resíduos e o centro dos resultados para a construção dos modelos PLS, SPA-MLR e iSPA-PLS, respectivamente. (d-f) Etapa de otimização do número de variáveis latentes no PLS, das variáveis individuais no SPA-MLR, e dos intervalos no iSPA-PLS, respectivamente. (g-i) Gráficos de valores previstos versus valores de referência, que incluem as amostras de calibração (círculos verdes) e as amostras de predição (quadrados verdes).....	48
Figura 4 – Valores de VIP-score para cada variável incluída no modelo PLS para a determinação do teor de gordura em amostras de leite de cabra. A linha preta contínua representa o limite típico, enquanto as linhas verticais verdes indicam os valores de VIP-score	49
Figura 5 – Aspectos relacionados à determinação da adulteração com leite de vaca em amostras de leite de cabra utilizando os espectros NIR pré-processados com suavização Savitzky-Golay e janela de 11 pontos (SGS): (a-c) Gráficos de influência que mostram as distâncias dos resíduos ao centro dos resultados para a construção dos modelos PLS, SPA-MLR e iSPA-PLS, respectivamente. (d-f) Etapa de otimização do número de variáveis latentes no PLS, das variáveis individuais no SPA-MLR, e dos intervalos no iSPA PLS. (g-i) Gráficos de valores previstos versus valores de referência, mostrando a distribuição das amostras de calibração (círculos roxos) e de predição (quadrados roxos)	53
Figura 6 – Valores de VIP score para cada variável incluída no modelo PLS para a	

determinação de leite de vaca em amostras de leite de cabra. A linha preta contínua representa o limite típico, enquanto as linhas verticais roxas indicam os valores de VIP score ...

.....54

LISTA DE TABELAS

	Página
Tabela 1. Composição básica dos leites caprino, bovino e humano (média por 100 g).	21
Tabela 2. Principais constituintes das proteínas do leite de cabra e suas composições..	23
Tabela 3. Constituintes dos lipídeos do leite de cabra e suas composições	24
Tabela 4. Resultados obtidos para a quantificação do teor de gordura nas amostras de leite de cabra e suas misturas com leite de vaca utilizando PLS, iSPA-PLS e SPA-MLR.....	46
Tabela 5: Resultados obtidos para a quantificação de leite de vaca no leite de cabra usando PLS, iSPA-PLS e SPA-MLR.....	50

LISTA DE ABREVIATURAS E SIGLAS

- DD-SIMCA – Modelagem Independente e Flexível por Analogia de Classe Dirigida aos Dados (do inglês, *Data-Driven Soft Independent Modeling of Class Analogy*)
- iSPA-PLS-DA – Algoritmo de Projeções Sucessivas para seleção de intervalos na Análise Discriminante por Mínimos Quadrados Parciais (do inglês, *the Successive Projections Algorithm for interval selection in Partial Least Squares Discriminant Analysis*)
- MFGM – Membrana do Glóbulo de Gordura do Leite (do inglês, *Milk Fat Globule Membrane*)
- MIR – Infravermelho médio (do inglês, *mid infrared*)
- MLR – Regressão Linear Múltipla (do inglês *Multiple Linear Regression*)
- MSC – Correção Multiplicativa de Espalhamento (do inglês, *Multiplicative Scatter Correction*)
- NIR – Infravermelho próximo (do inglês, *near infrared*)
- OFF – Correção de Offset (do inglês, *Offset Correction*)
- PLS – Mínimos Quadrados Parciais (do inglês, *Partial Least Squares*)
- REP – Erro Relativo de Predição (do inglês, *Relative Error of Prediction*)
- RMSECV – Erro Quadrático Médio de Validação Cruzada (do inglês, *Root Mean Square Error of Cross-Validation*)
- RMSEP – Erro Quadrático Médio de Predição (do inglês, *Root Mean Square Error of Prediction*)
- RMSEV – Erro Quadrático Médio de Validação (do inglês, *Root Mean Square Error of Validation*)
- RPD – Razão de Desempenho para Desvio (do inglês, *Ratio of Performance to Deviation*)
- SGS – Suavização de Savitzky-Golay (do inglês, *Savitzky-Golay Smoothing*)
- SNV – Variação Normal Padrão (do inglês, *Standard Normal Variate*)
- SPA – Algoritmo das Projeções Sucessivas (do inglês *Successive Projections Algorithm*)
- VIP – Importância das Variáveis na Projeção (do inglês, *Variable Importance in Projection*)

SUMÁRIO

1 INTRODUÇÃO.....	15
2 OBJETIVOS.....	19
2.1 Objetivo Geral.....	19
2.2 Objetivos Específicos	19
3 REFERENCIAL TEÓRICO	20
3.1 Leite de cabra	20
3.1.1 Fração Proteica	22
3.1.2 Fração Lipídica	23
3.2 Espectrometria de Infravermelho Próximo (NIR)	25
3.3 Quimiometria.....	27
3.3.1 Pré-processamento de Dados	27
3.3.1.1 Correção de offset (Offset Correction – OFF)	28
3.3.1.2 Correção linear de linha de base (Linear Baseline Correction – LBC)	28
3.3.1.3 Correção multiplicativa de espalhamento (Multiplicative Scatter Correction – MSC)	28
3.3.1.4 Variação normal padrão (Standard Normal Variate – SNV).....	29
3.3.1.5 Suavização de Savitzky-Golay (Savitzky-Golay Smoothing – SGS)	29
3.3.1.6 Derivadas.....	30
3.4 Calibração Multivariada.....	30
3.4.1 Mínimos Quadrados Parciais (Partial Least Squares - PLS)	31
3.4.1.1 Fase de Calibração	31
3.4.1.2 Fase de Validação	32
3.4.1.3 Determinação do Número Ótimo de Variáveis Latentes.....	33
3.4.2 Regressão Linear Múltipla (Multiple Linear Regression-MLR)	34
3.4.3 Algoritmo de Projeções Sucessivas (Successive Projections Algorithm – SPA)	35
4 METODOLOGIA	38
4.1 Amostras.....	38
4.2 Medições por NIR.....	38
4.3 Procedimento quimiométrico	40
5 RESULTADOS E DISCUSSÃO	42
5.1 Investigação espectral.....	42
5.2 Quantificação do teor de gordura	43

5.3 Quantificação do teor de adulteração com leite de vaca.....	49
6 CONCLUSÕES	56
REFERÊNCIAS.....	57

1 INTRODUÇÃO

O leite de cabra se destaca por seu perfil nutricional rico e propriedades únicas, incluindo menor alergenicidade e melhor digestibilidade, quando comparadas ao leite de vaca. Essas características distintas são atribuídas principalmente à sua composição química específica. A alergenicidade do leite de cabra é influenciada pelo polimorfismo genético da α_{S1} -caseína, sendo que uma menor imunorreatividade está associada aos genótipos nulos de α_{S1} -caseína (0101 ou 01F). Em contraste, a presença de α_{S1} -caseína e β -lactoglobulina são as principais proteínas responsáveis pela alergia às proteínas do leite de vaca (Verruck, Dantas & Prudencio, 2019; Pereira *et al.*, 2022). Particularmente, as proteínas e os lipídios do leite de cabra apresentam maior digestibilidade em comparação com os de outras espécies em condições semelhantes. Essa digestibilidade aumentada está relacionada aos glóbulos de gordura menores, que facilitam o metabolismo lipídico humano (Liao *et al.*, 2024). Além disso, o leite de cabra é caracterizado por uma maior concentração de ácidos graxos de cadeia curta, que são mais facilmente digeríveis. O teor de gordura é notavelmente variável, tanto em termos qualitativos quanto quantitativos, e pode ser influenciado por fatores como dieta do animal, práticas de manejo, estágio de lactação, raça, genótipo e até mesmo níveis de estresse (Nayik *et al.*, 2022).

A preferência pelo leite de cabra é frequentemente motivada por restrições alimentares ou considerações de saúde, o que reforça ainda mais a importância da sua integridade. Os benefícios nutricionais do leite de cabra e fatores como a disponibilidade geográfica e a sazonalidade contribuem para seu elevado valor comercial. No mercado global de laticínios, o leite de cabra e seus derivados são altamente valorizados e procurados por consumidores que buscam alternativas ao leite de vaca, mais comum e menos dispendioso. Esse maior valor de mercado, no entanto, torna o leite de cabra vulnerável a fraudes e adulterações, o que pode afetar negativamente a qualidade do produto e a saúde do consumidor (Clark & García, 2017).

A fraude no leite envolve a adição ou remoção intencional de componentes valiosos, levando à adulteração com motivação econômica que compromete a integridade do produto, reduz o valor nutricional e representa potenciais riscos à saúde dos consumidores (Moore, Spink & Lipp, 2012). No caso do leite de cabra, a adulteração geralmente envolve a adição de substâncias de menor custo, como água, soro ou leite de outras espécies, principalmente bovina, devido à sua abundância e menor valor de mercado. No entanto, a detecção de adulterações em líquidos lácteos é desafiadora devido à natureza complexa e similar das matrizes. Variações significativas na composição de macronutrientes podem ocorrer dentro da mesma espécie, influenciadas por fatores ambientais e genéticos (Patil *et al.*, 2024).

Compreender a variação normal dos macronutrientes é essencial para interpretar corretamente os dados analíticos e distinguir variabilidade natural de adulterações. Na ciência dos alimentos, considerar essas variações permite o estabelecimento de referências nutricionais confiáveis, garantindo a qualidade do produto. Isso é particularmente importante no desenvolvimento de modelos espectroscópicos baseados em Quimiometria, aumentando a robustez e a capacidade preditiva, ao mesmo tempo em que minimiza o risco de classificações incorretas ou conclusões errôneas.

Embora existam diversos métodos bem estabelecidos para detectar adulterações no leite, incluindo técnicas baseadas em DNA, eletroforese, cromatografia e ensaios imunoenzimáticos, esses métodos geralmente exigem preparo extensivo da amostra, tempo considerável, instrumentação avançada e uso substancial de reagentes químicos (Patil *et al.*, 2024; Pereira *et al.*, 2022; Clark & García, 2017).

Nesse contexto, metodologias mais simples baseadas nos princípios da química verde têm ganhado destaque por sua eficiência, custo reduzido e pouca ou nenhuma necessidade de preparo prévio da amostra (Pereira *et al.*, 2020). Entre elas, as técnicas espectroscópicas vibracionais, incluindo espectroscopia Raman (Li *et al.*, 2023; Batesttin *et al.*, 2022; Yazgan *et al.*, 2020; Yaman, 2020), de infravermelho médio (MIR) (Du *et al.*, 2025; Chu *et al.*, 2023; Sen *et al.*, 2021; Yaman, 2020) e de infravermelho próximo (NIR) (He *et al.*, 2023; Pereira *et al.*, 2021, 2020; Teixeira *et al.*, 2020), vêm sendo recentemente exploradas para abordar o problema da adulteração do leite de cabra. Elas também têm sido aplicadas em produtos derivados de leite de cabra, como leite em pó, fórmulas infantis, bebidas, iogurtes e queijos (Ferreira *et al.*, 2025; Zhao *et al.*, 2022; Teixeira *et al.*, 2022, 2021). Entre esses métodos não destrutivos e vantajosos, os espectrômetros NIR miniaturizados se destacam por serem de baixo custo, fáceis de usar e portáteis. Embora instrumentos compactos possuam faixas espectrais menores e menores razões sinal-ruído, estudos demonstraram diferenças mínimas ou não significativas na classificação ou quantificação para a maioria das aplicações analíticas (Foli *et al.*, 2024).

Pereira *et al.* (2020) desenvolveram uma metodologia para quantificação de gordura e proteína no leite de cabra e detecção de adulteração com leite de vaca utilizando um instrumento NIR de bancada combinado com regressão por Mínimos Quadrados Parciais (PLS), seleção de intervalos em PLS (iPLS) e o Algoritmo de Projeções Sucessivas para seleção de intervalos em PLS (iSPA-PLS). Embora essa abordagem tenha representado um avanço significativo em relação aos métodos tradicionais para determinação do teor de gordura e proteína e detecção de adulterações, ainda permanecia inacessível para pequenos produtores e cooperativas. O alto custo dos instrumentos NIR de bancada e a necessidade de um ambiente laboratorial dedicado

limitam sua aplicação prática em campo. Para superar essa limitação, Pereira *et al.* (2021) desenvolveram uma metodologia mais acessível para identificar se as amostras de leite de cabra estavam adulteradas com leite de vaca, utilizando um espectrômetro NIR portátil acoplado ao algoritmo iSPA-PLS adaptado para Análise Discriminante (iSPA-PLS-DA). Isso representou um avanço em termos de acessibilidade metodológica, pois os dispositivos NIR portáteis são mais acessíveis e podem ser utilizados fora do ambiente laboratorial. No entanto, essa abordagem permitia apenas a detecção da fraude, sem quantificar o grau de adulteração ou determinar os teores de proteína e gordura. Esses parâmetros são essenciais para avaliar a qualidade nutricional do leite de cabra, garantir a segurança de indivíduos com alergia ao leite de vaca e monitorar possíveis alterações no bem-estar animal e nas práticas de manejo.

Os resultados de Pereira *et al.* (2020, 2021) destacam os benefícios do uso de abordagens de seleção de variáveis para aprimorar metodologias de análise de alimentos. Nesse sentido, o Algoritmo de Projeções Sucessivas (SPA) se destaca como uma alternativa robusta para enfrentar desafios analíticos. O SPA é uma técnica determinística de seleção de variáveis, projetada especificamente para reduzir a multicolinearidade e eliminar variáveis redundantes por meio de projeções vetoriais, garantindo assim uma solução única durante o processo de seleção (Gomes *et al.*, 2022). Além de sua aplicação em tarefas de classificação, o SPA melhora significativamente o desempenho de modelos em calibração multivariada, especialmente quando integrado a algoritmos de Regressão Linear Múltipla (SPA-MLR) e Mínimos Quadrados Parciais (iSPA-PLS) (Vallese *et al.*, 2024; Gomes *et al.*, 2022).

O SPA-MLR opera em três fases: (1) geração de cadeias de variáveis minimamente redundantes por meio de projeções sucessivas, de forma independente da variável dependente y ; (2) seleção do melhor subconjunto com base em uma função objetivo, como o erro quadrático médio de validação (RMSEV) ou de validação cruzada (RMSECV); e (3) refinamento do subconjunto pela eliminação de variáveis irrelevantes com base nos coeficientes de regressão. Já o iSPA-PLS estende o SPA dividindo o X_{cal} em intervalos de variáveis, representando cada intervalo pela coluna de maior norma. Esses intervalos informam as projeções e permitem a modelagem por PLS. Os intervalos ótimos e o número de variáveis latentes são selecionados com base nos valores de RMSEV/RMSECV (Vallese *et al.*, 2024).

Ao melhor do nosso conhecimento, a aplicação de dispositivos portáteis de espectroscopia NIR para quantificação da adulteração do leite de cabra com leite de vaca ainda não foi reportada, apesar de Musa & Yang (2021) já terem demonstrado a viabilidade do uso de um dispositivo NIR miniaturizado para quantificar diversos adulterantes (água, ureia, amido e leite de cabra) em leite de vaca. Portanto, o estudo proposto é pioneiro na utilização de um

espectrômetro NIR miniaturizado, integrado com calibração multivariada baseada em SPA por meio de algoritmos de seleção de variáveis e intervalos (especificamente, SPA-MLR e iSPA-PLS), para determinar o teor de gordura e a adulteração com leite de vaca em amostras de leite de cabra. Esse avanço pode melhorar significativamente a viabilidade do controle de qualidade em tempo real para pequenos produtores e cooperativas, abordando lacunas críticas na área e promovendo a democratização dos métodos analíticos.

2 OBJETIVOS

2.1 Objetivo Geral

Desenvolver uma nova metodologia analítica verde e de baixo custo para a quantificação de adulterações em leite de cabra fluido empregando um espectrofotômetro NIR portátil e técnicas quimiométricas de calibração multivariada.

2.2 Objetivos Específicos

- Quantificar o teor de lipídios em leite de cabra fluido empregando um espectrofotômetro NIR portátil e calibração multivariada;
- Quantificar o teor de adulteração por adição de leite de vaca em leite de cabra fluido empregando um espectrofotômetro NIR portátil e calibração multivariada;
- Comparar o desempenho dos modelos quimiométricos baseados em seleção de variáveis pelo Algoritmo das Projeções Sucessivas frente aos modelos baseados em informação espectral completa.

3 REFERENCIAL TEÓRICO

3.1 Leite de cabra

De acordo com a Instrução Normativa nº 51, de 18 de setembro de 2002, do Ministério da Agricultura, Pecuária e Abastecimento (BRASIL, Ministério da Agricultura, Pecuária e Abastecimento, 2002), leite é o produto proveniente da ordenha completa e ininterrupta de vacas sadias, realizada sob condições higiênicas. Para leite de outras espécies, este deve ser identificado especificamente, como no caso do leite caprino, proveniente da ordenha de animais da espécie caprina (BRASIL, 2000).

Na biologia, o leite é definido como o produto resultante da secreção das glândulas mamárias de fêmeas mamíferas, com a função primordial de alimentar o recém-nascido, fornecendo energia através de lipídios, lactose e proteínas, além de aminoácidos essenciais, ácidos graxos, vitaminas, minerais e água. As exigências nutricionais dos recém-nascidos variam de acordo com a maturidade, taxa de crescimento e demandas energéticas, que são influenciadas por fatores ambientais. Consequentemente, a composição do leite difere significativamente entre espécies, refletindo essas necessidades. Além disso, o leite desempenha funções fisiológicas cruciais, principalmente por meio de proteínas e peptídeos, como imunoglobulinas, enzimas, inibidores de enzimas, fatores de crescimento, hormônios e agentes antibacterianos, apoiando a função imunológica, o crescimento e o desenvolvimento (O'mahony e Fox, 2014; Ordóñez, 2005).

Físico-quimicamente, trata-se de uma mistura heterogênea composta por substâncias como lactose, glicérides, proteínas, sais, vitaminas e enzimas, que se apresentam em diferentes estados: em emulsão (gordura e vitaminas lipossolúveis), suspensão (caseínas ligadas a sais minerais) e dissolução (lactose, vitaminas hidrossolúveis, proteínas do soro, sais e outras substâncias) (Ordóñez, 2005).

O leite é uma solução aquosa (soro do leite) contendo lactose, sais inorgânicos e orgânicos, além de compostos em níveis traço, com partículas coloidais em três faixas: proteínas do soro dissolvidas em nível molecular, caseínas dispersas em micelas (50–500 nm) e lipídios emulsificados em glóbulos (1–20 µm). A estabilidade coloidal das micelas de caseína é essencial nutricional e tecnologicamente. Elas se desestabilizam e se agregam ou gelificam após proteólise limitada ou acidificação (pH ~4,6). *In vivo*, isso ocorre no estômago do recém-nascido, retardando o trânsito e melhorando a digestão, enquanto tecnologicamente pode ser controlado para produção de queijos e lácteos fermentados (O'mahony e Fox, 2014).

A **Tabela 1** apresenta uma comparação entre a composição básica dos leites caprino,

bovino e humano (valores médios por 100 g).

Tabela 1. Composição básica dos leites caprino, bovino e humano (média por 100 g).

	Caprino	Bovino	Humano
Composição			
Sólidos totais (g)	12,2	12,3	12,3
Lipídeos (%)	4,0-4,5	3,8	4,1
Proteínas (%)	3,2	3,3	1,3
Lactose (%)	4,6	4,7	7,2
Cinzas (g)	0,8	0,7	0,2
Água (%)	87,5	87,7	86,7
Energia (kcal)	70	69	68
Elementos (mg/100g)			
Na	34,0	50,0	14,0
K	180,0	150,0	58,0
Ca	129,0	120,0	34,0
Mg	20,0	12,0	3,0
P	106,0	95,0	14,0
Fe	0,04-0,1	0,05	0,07
Cl	130,0	95,0	42,0
Vitaminas (por 100g)			
Vitamina A (IU)	185,0	126,0	241,0
Tiamina (mg)	0,05	0,04	0,014
Riboflavina (mg)	0,14	0,16	0,04
Ácido pantotênico (mg)	0,31	0,314	-
Niacina (mg)	0,28	0,08	0,18
Vitamina B (mg)	0,05	0,04	0,01
Ácido fólico (mg/L)	6,0	50,0	56,0
Vitamina B ₁₂ (mg)	0,05	0,14	0,14
Biotina (mg)	2,00	2,00	0,70
Vitamina C (mg)	1,50	1,50	1,00
Vitamina D (mg)	0,06	0,03	0,025

Fonte: Yadav, Singh e Yadav (2016).

Dentro de uma espécie, a composição do leite varia dependendo de fatores como raça,

dieta, ambiente e estágio de lactação (Silanikove *et al.*, 2010). As concentrações dos constituintes principais variam amplamente entre as espécies: lipídios, 2–55%; proteínas, 1–20%; e lactose, 0–10%, refletindo principalmente as necessidades energéticas (lipídios e lactose) e a taxa de crescimento (principalmente proteínas) do recém-nascido. As concentrações dos constituintes minoritários também variam significativamente (por exemplo, lisozima e lactoferrina em leites equino, humano e bovino). O teor de gordura do leite bovino apresenta grandes variações entre raças, e dentro de uma mesma raça há ampla variação no teor de gordura e proteínas entre animais individuais; diferenças semelhantes ocorrem no leite de ovelhas, cabras e búfalos (O'mahony e Fox, 2014).

No caso do leite caprino, é fortemente influenciado pelo estado de lactação, pois os constituintes são mais elevados no início da fase de produção do leite, diminuem rapidamente, permanecem baixos por um tempo e aumentam novamente nos estágios finais dessa produção (Silanikove *et al.*, 2010). Além disso, nutricionalmente, o leite caprino é considerado excelente devido às suas propriedades, sendo rico em vitaminas, como a vitamina A, e minerais, como cálcio, fósforo, potássio e magnésio. Apresenta também elevado valor proteico e é recomendado para pessoas com restrições alimentares, pois seus glóbulos de gordura são menores, o que resulta em melhor digestibilidade comparado ao leite bovino (Pinheiro, 2012).

3.1.1 Fração Proteica

A fração proteica do leite de ruminantes é composta principalmente por caseínas e proteínas do soro. As caseínas, que formam a fase micelar, representam cerca de 80% das proteínas, e são divididas em quatro tipos: α_{s1} -caseína, α_{s2} -caseína, β -caseína e κ -caseína. As proteínas do soro compõem a fase solúvel (Singh, Kumar e Singh, 2022; Pereira, 2020).

No leite de cabra, há uma menor concentração de α_s -caseínas e maiores quantidades de β -caseína em comparação ao leite bovino. Além disso, as micelas de caseína no leite caprino contêm mais cálcio e fósforo e têm menor estabilidade térmica. Essas diferenças são importantes na distinção entre os leites de cabra e vaca (Singh, Kumar e Singh, 2022). Os seres humanos podem digerir de forma eficaz a lactose e a β -lactoglobulina presentes no leite de cabra, uma vez que ele contém menor concentração de lactose e α_{s1} -caseína. Como resultado, o leite de cabra é menos alergênico do que o leite de vaca (Nayik *et al.*, 2022). Os diferentes constituintes das proteínas do leite de cabra e suas composições estão apresentados na **Tabela 2**.

Tabela 2. Principais constituintes das proteínas do leite de cabra e suas composições.

Constituintes	Concentração	Unidade
Proteínas majoritárias		
α_s -caseína	26	(% de caseínas totais)
β -caseína	64	
κ -caseína	10	
α_s -caseína/ β -caseína	0,41 (taxa em %)	
Lactoferrina	20–200	$\mu\text{g/mL}$
Transferrina	20–200	
Proteína de ligação ao folato	12	
Prolactina	44	ng/mL
Ribonuclease	425	$\mu\text{g/dL}$
Lisozima	25	
Lipase	36	$\mu\text{mol/dL}$
Desidrogenase málica	50	$\mu\text{mol/s por mL}$
Desidrogenase láctica	47	
Xantina oxidase	19–113	$\mu\text{L O}_2/\text{h}$
Fosfatase alcalina	11–13	mg/L
Carnitina	16,4	
Ácido orótico	13	
Aminoácidos livres	48	
ATP	19	
Protease de ácido siálico	13,89	
Elementos (mg/100g)		
IgA	30–80	$\mu\text{g/mL}$ (leite)
	0,9–2,4	mg/mL (coloostro)
IgM	10–40	$\mu\text{g/mL}$ (leite)
	1,6–5,2	mg/mL (coloostro)
IgG	100–400	$\mu\text{g/mL}$ (leite)
	50–60	mg/mL (coloostro)

Fonte: Nayik *et al.* (2022).

3.1.2 Fração Lipídica

A fração lipídica é a componente mais variável da matéria seca no leite de diferentes espécies, contribuindo significativamente para as variações intra e interespecíficas (Pereira, 2021). O leite caprino se distingue por apresentar glóbulos de gordura menores em comparação ao leite bovino, o que resulta em uma emulsão mais homogênea e facilita a digestibilidade e

absorção dos lipídios (Silanikove *et al.*, 2010).

Adicionalmente, o leite caprino contém uma maior proporção de ácidos graxos de cadeia curta e média, que são metabolizados mais facilmente. Em contrapartida, o leite bovino apresenta um maior percentual de ácidos graxos de cadeia longa. Essa diferença está relacionada à melhor digestibilidade e aos efeitos nutricionais mais favoráveis do leite de cabra no organismo (Silanikove *et al.*, 2010). Os diversos constituintes dos lipídeos presentes no leite de cabra e suas composições estão apresentados na **Tabela 3**.

Tabela 3. Constituintes dos lipídeos do leite de cabra e suas composições.

Componente	Componentes-chave	Concentração (%)
Ácido graxo saturado	Butírico (C4:0)	2,18
	Capróico (C6:0)	2,39
	Caprílico (C8:0)	2,73
	Cáprico (C10:0)	9,97
	Láurico (C12:0)	4,99
	Mirístico (C14:0)	9,81
	Palmítico (C16:0)	28,00
	Estearico (C18:0)	8,88
	Pendadecanóico (C15:0)	0,71
Ácido graxo monoinsaturado	Miristoleico (C14:1)	0,18
	Palmitoleico (C16:1)	1,59
	Oleico (C18:1)	19,3
Ácido graxo poli-insaturado	Linoleico (C18:2)	3,19
	Linolênico (C18:3)	0,42
	Linoleico conjugado (C18:2)	0,70
Ácidos graxos minoritários	-	3,19
Fração fosfolipídica (%)	Fosfatidil etanolamina	25,4
	Fosfatidil colina	28,2
	Esfingomielina	29,2
	Fosfatidil Inositol	4,0
	Fosfatidil serina	3,2

Fonte: Nayik *et al.* (2022).

3.2 Espectrometria de Infravermelho Próximo (NIR)

A espectroscopia no infravermelho é uma técnica amplamente utilizada na análise química, devido à sua capacidade de fornecer informações detalhadas sobre vibrações moleculares e ligações químicas presentes em diferentes substâncias. Essa técnica baseia-se na interação da radiação infravermelha com a matéria, permitindo a identificação de compostos e o monitoramento de mudanças estruturais (Pasquini, 2003; Fodor *et al.* 2024). O espectro infravermelho divide-se em três regiões principais: o infravermelho próximo (*Near Infrared* - NIR), o infravermelho médio (*Mid Infrared* - MIR) e o infravermelho distante (*Far Infrared* - FAR), cada uma com características específicas que ampliam a versatilidade da espectroscopia na investigação de materiais orgânicos e inorgânicos (Skoog *et al.*, 2009).

Graças a essas características, a espectroscopia tem se tornado uma ferramenta analítica de destaque na indústria alimentícia, especialmente na indústria láctea, onde tem sido amplamente aplicada. Técnicas espectroscópicas vibracionais, como as espectroscopias Raman (Li *et al.*, 2023; Batestin *et al.*, 2022; Yazgan *et al.*, 2020; Yaman, 2020), MIR (Du *et al.*, 2025; Chu *et al.*, 2023; Sen *et al.*, 2021; Yaman, 2020) e NIR (He *et al.*, 2023; Pereira *et al.*, 2021; Teixeira *et al.*, 2020), têm se destacado no estudo de adulterações em leite de cabra e seus derivados, como queijos, iogurtes, fórmulas infantis e leite em pó (Ferreira *et al.*, 2025; Zhao *et al.*, 2022; Teixeira *et al.*, 2022).

Entre essas técnicas, a espectroscopia de infravermelho próximo (NIR) destaca-se por sua eficiência analítica na faixa de 13.300 cm^{-1} a 4.000 cm^{-1} , correspondente a comprimentos de onda entre 750 e 2.500 nm. Essa abordagem mede a energia absorvida por moléculas e permite obter informações qualitativas e quantitativas das amostras, resultantes da interação entre os constituintes moleculares e as ondas NIR (Pasquini, 2003). Na região NIR, a absorção ocorre em ligações moleculares, como O-H, C-H, C-O e N-H, que vibram ao interagir com a radiação, gerando espectros de absorção específicos (Cen; He, 2007).

A aplicação da NIR é amplamente reconhecida por suas vantagens, como rapidez, caráter não destrutivo, portabilidade e mínima necessidade de preparo das amostras, necessitar de menor especialidade técnica, gerar uma quantidade menor de resíduos, e demandar menos energia quando comparada a métodos tradicionais (Pasquini, 2018; Fodor *et al.* 2024).

Na indústria láctea, a técnica é utilizada para avaliar a composição química do leite, incluindo teores de gordura, proteína e lactose, além de detectar adulterações, como a adição de leite de outras espécies ou componentes não desejados. Por meio de padrões espectrais específicos, é possível identificar fraudes e garantir a conformidade do produto com padrões regulatórios, promovendo segurança alimentar e qualidade (Pereira *et al.* 2021).

Apesar das vantagens, a técnica apresenta limitações, como baixa intensidade de radiação, o que reduz sua capacidade de detectar traços em concentrações extremamente baixas. O NIR também não é um método absoluto, uma vez que depende do método de referência utilizado (Pasquini, 2018; Fodor *et al.* 2024). Além disso, fatores como variações nos equipamentos, condições ambientais e preparo inadequado das amostras podem gerar ruídos que dificultam a análise, principalmente em matrizes complexas, como é o caso dos alimentos, incluindo-se aqui o leite de cabra (Pereira, 2020). No entanto, avanços em Quimiometria têm ampliado a aplicabilidade do NIR, permitindo a seleção de variáveis e a padronização das análises, aspectos cruciais para a investigação de matrizes alimentares complexas (Cen; He, 2007; Fodor *et al.* 2024).

Uma forma de contornar o desafio da avaliação de espectros obtidos no infravermelho próximo (NIR), devido à sobreposição de bandas resultantes de combinações e harmônicos das vibrações moleculares, é o uso de técnicas de pré-processamento. Dentre essas técnicas, a correção de espalhamento multiplicativo (*multiplicative scattering correction* – MSC), a variação normal padrão (*standard normal variate* – SNV) e os algoritmos de suavização, como Savitzky-Golay, sendo estes amplamente empregadas, e fornecendo melhores soluções para esses desafios. Tais métodos não apenas melhoram a resolução espectral, mas também corrigem desvios na linha de base, permitindo maior precisão em análises quantitativas, especialmente em combinações com técnicas quimiométricas, como a Análise de Componentes Principais (*Principal Component Analysis* – PCA) e a Regressão por Mínimos Quadrados Parciais (*Partial Least Squares Regression* – PLS) (Fodor *et al.* 2024). É importante, no entanto, ressaltar que ao se aplicar pré-processamentos, é fundamental a escolha cuidadosa da melhor técnica, para que não haja perda de informações valiosas nos dados espectrais. Isso influencia diretamente na robustez dos modelos analíticos, garantindo precisão e confiabilidade nas análises alimentares, especialmente naquelas que utilizam espectroscopia NIR (Rinnan; Van Den Berg; Engelsen, 2009)

Estudos recentes reforçam o potencial da espectroscopia NIR para análise do leite de cabra. Pereira *et al.* 2021 aplicaram o Algoritmo das Projeções Sucessivas para seleção de intervalos em Análise Discriminante por Mínimos Quadrados Parciais (*Interval selection by the Successive Projections Algorithm in Partial Least Squares Discriminant Analysis* – iSPA-PLS-DA) com um dispositivo NIR portátil para identificar adulterações em leite de cabra com leite de vaca, enquanto (Musa e Yang, 2021) usaram a Modelagem Independente e Flexível por Analogia de Classe Orientada aos Dados (*Data-Driven Soft Independent Modeling of Class Analogy* – DD-SIMCA) e PLS para detectar e quantificar adulterantes em leite de vaca. Ambos

os trabalhos apontam resultados satisfatórios. Contudo, a quantificação específica de adulterações em leite de cabra ainda exige pesquisas mais abrangentes.

3.3 Quimiometria

Ao se interpretar os espectros obtidos por espectroscopia NIR, fatores como transmitância, refletância, tamanho das partículas, posição da amostra, dentre outros, podem influenciar diretamente na geração de espectros. Isto pode, por vezes, prejudicar a elaboração de modelos preditivos, pois esses espectros podem ser amplamente influenciados por não-linearidades adicionadas pela dispersão da luz (Rinnan; Van Den Berg; Engelsen, 2009). Essa constatação destaca a necessidade de utilizar ferramentas que minimizem influências indesejadas, tornando assim as análises com NIR um terreno fértil para o surgimento de uma gama de pré-processamentos, especialmente quando se analisa amostras complexas e multiconstituintes como as amostras biológicas, onde o tamanho das partículas e o comprimento de radiação eletromagnética podem ser comparados (Rinnan; Van Den Berg; Engelsen, 2009). Mediante isso, estratégias de pré-tratamentos de dados para reduzir o ruído se mostram interessantes para extrair informações relevantes da espectroscopia no NIR, particularmente quando são empregados dispositivos miniaturizados (Bêc; Grabska; Huck, 2022).

3.3.1 Pré-processamento de Dados

Após a obtenção e organização, os dados devem ser submetidos, quando necessário, a técnicas de pré-processamento, cuja finalidade principal é reduzir ruídos indesejáveis, como o ruído instrumental, os efeitos de espalhamento de luz ou variações na espessura da amostra. Esse tipo de abordagem é comumente empregado em análises multivariadas de dados espectrais, sendo fundamental para melhorar a qualidade dos modelos preditivos. No entanto, sua aplicação requer cautela, pois, se mal conduzida, pode causar distorções, perda de informações relevantes e até mesmo comprometer a validade dos resultados obtidos (Ferreira, 2015).

Diversas técnicas de pré-processamento podem ser testadas ao longo da análise a fim de se identificar aquelas que melhor contribuem para a robustez e desempenho dos modelos. Esses tratamentos podem ser empregados tanto nas amostras como nas variáveis da matriz de dados (Ferreira, 2015).

Entre as técnicas de pré-processamento algumas merecem destaque devido à sua ampla utilização e eficácia na correção de interferências específicas, entre os quais minimizar

variações aleatórias, como o ruído experimental, bem como remover informações sistemáticas indesejáveis. O ruído instrumental, por exemplo, pode ser atenuado por meio de técnicas de alisamento, como a suavização por média móvel. Já as variações sistemáticas podem ser corrigidas ou reduzidas utilizando métodos como derivadas de Savitzky-Golay, variação normal padrão (SNV), correção de espalhamento multiplicativo (MSC) e correção de linha de base (Ferreira, 2015).

3.3.1.1 Correção de offset (Offset Correction – OFF)

Algumas técnicas de pré-processamento, como o MSC e o SNV, incluem em sua operação a correção de offset, que visa eliminar variações aditivas constantes nos espectros. Tais variações podem resultar de ruídos instrumentais ou instabilidades nas condições de medição. Segundo Rinnan; Van Den Berg; Engelsen (2009), a remoção desses efeitos não químicos, como deslocamentos da linha de base e inclinações espectrais (*slope*), é essencial para reduzir a influência de fenômenos de espalhamento da luz, comuns em espectroscopia NIR. A correta aplicação desses métodos contribui para o alinhamento dos espectros e favorece a construção de modelos preditivos mais robustos e menos propensos ao sobreajuste.

3.3.1.2 Correção linear de linha de base (Linear Baseline Correction – LBC)

Correção da linha de base é uma etapa essencial no pré-processamento de espectros NIR, voltada à remoção de variações sistemáticas lentas e não informativas que ocorrem ao longo do eixo espectral. Tais variações, que podem surgir por interferências físicas, instrumentais ou relacionadas à preparação da amostra, manifestam-se como deslocamentos ou curvaturas que não guardam relação direta com a composição química do material analisado. Segundo Rinnan; Van Den Berg; Engelsen (2009), essas alterações podem ser corrigidas por meio de ajustes polinomiais, geralmente de segunda ordem, aplicados ao longo dos comprimentos de onda. Essa abordagem, conhecida como detrending ou correção de linha de base, visa subtrair do espectro original uma tendência sistemática, preservando apenas as variações relevantes para a análise química. A aplicação adequada desse tipo de correção contribui para reduzir a complexidade dos modelos multivariados e aumentar sua robustez preditiva.

3.3.1.3 Correção multiplicativa de espalhamento (Multiplicative Scatter Correction – MSC)

A correção de espalhamento multiplicativo (MSC) é uma técnica de pré-processamento amplamente utilizada para remover efeitos de espalhamento aditivos e multiplicativos que afetam a absorbância em espectros. Tais efeitos podem ser causados por fatores físicos diversos, como variações no caminho óptico, na sensibilidade dos detectores, na temperatura, na pressão e, especialmente, nas características físicas das amostras, como tamanho e forma das partículas em amostras sólidas, ou ainda em emulsões e dispersões. Esses fatores introduzem variabilidade nos dados espectrais que não está relacionada à composição química da amostra e, portanto, sua eliminação melhora a qualidade dos modelos multivariados (Ferreira, 2015).

De forma simplificada, a MSC projeta cada espectro individual sobre um espectro de referência, tipicamente o espectro médio do conjunto, por meio de uma regressão linear simples. Esse procedimento corrige as diferenças sistemáticas causadas por variações físicas, como o tamanho das partículas, que afetam o caminho óptico e outras propriedades das amostras (Rinnan; Van Den Berg; Engelsen, 2009).

3.3.1.4 Variação normal padrão (Standard Normal Variate – SNV)

A variação normal padrão (SNV) é uma técnica de pré-processamento amplamente utilizada na análise de espectros, reconhecida por sua eficácia na correção de efeitos de espalhamento de luz e variações na densidade e no tamanho das partículas sólidas presentes na amostra. Embora matematicamente semelhante à MSC ao atuar na remoção de efeitos aditivos e multiplicativos, a SNV difere no método, pois corrige cada espectro individualmente sem a necessidade de um espectro de referência. No SNV, cada espectro é centrado pela sua média e normalizado pelo seu desvio padrão, equivalente a um autoescalamento linha a linha da matriz espectral. Como resultado, os espectros apresentam médias próximas de zero e variâncias unitárias, favorecendo análises multivariadas subsequentes, como PCA ou a regressão PLS. A combinação da SNV com outras técnicas, como derivadas ou alisamentos, pode potencializar a remoção de variabilidades espúrias, otimizando a qualidade dos dados espectrais (Ferreira, 2015; Rinnan, van den Berg & Engelsen, 2009).

3.3.1.5 Suavização de Savitzky-Golay (Savitzky-Golay Smoothing – SGS)

Após as correções globais de espalhamento, o método de Savitzky-Golay é amplamente utilizado para o alisamento dos sinais espectroscópicos, incluindo os espectros NIR, devido à sua capacidade de reduzir o ruído sem comprometer significativamente características

fundamentais do sinal, como picos e vales. Diferentemente da média móvel simples, o filtro de Savitzky-Golay ajusta um polinômio aos dados dentro de uma janela móvel, utilizando regressão por mínimos quadrados, e utiliza esse polinômio para estimar os valores suavizados (Ferreira, 2015).

Isso é feito de forma que para cada ponto do sinal, uma janela simétrica contendo $2m+1$ pontos é selecionada e um polinômio de grau n é ajustado. O valor central da janela é substituído pela avaliação desse polinômio naquele ponto. A janela se desloca sequencialmente ao longo do sinal, promovendo a suavização do espectro. A escolha do tamanho da janela e do grau do polinômio influencia diretamente a suavização: janelas maiores e polinômios de baixo grau suavizam mais o ruído, mas podem distorcer o sinal; janelas menores e polinômios de grau elevado preservam detalhes, mantendo parte do ruído (Rinnan, van den Berg & Engelsen, 2009; Ferreira, 2015).

Outro ponto a destacar é que o filtro de Savitzky-Golay é particularmente eficaz para o cálculo de derivadas, permitindo suavizar o sinal e obter suas derivadas simultaneamente, o que é essencial para a remoção de deslocamentos e inclinações na linha de base dos espectros simultaneamente (Ferreira, 2015).

3.3.1.6 Derivadas

No contexto espectral, a derivada de uma função representa a taxa de variação da absorbância em função do comprimento ou número de onda. A primeira derivada corrige deslocamentos constantes da linha de base, uma vez que a derivada de uma constante é zero, posicionando o espectro corrigido em torno de zero. Já a segunda derivada é eficaz para corrigir inclinações lineares na linha de base, comuns em espectros de refletância difusa, tornando o uso de pré-tratamentos derivados uma prática comum para melhorar a qualidade dos dados e a precisão das análises multivariadas (Ferreira, 2015; Rinnan, van den Berg & Engelsen, 2009).

3.4 Calibração Multivariada

Os métodos multivariados oferecem maior robustez analítica ao considerarem simultaneamente múltiplas variáveis, permitindo a construção de modelos mais representativos da realidade experimental. Ao utilizar diferentes respostas instrumentais de uma mesma amostra, esses métodos proporcionam melhor capacidade preditiva e interpretação de propriedades químicas (Ferreira, 2015).

3.4.1 Mínimos Quadrados Parciais (*Partial Least Squares - PLS*)

A regressão por mínimos quadrados parciais (PLS) é uma técnica estatística amplamente empregada em quimiometria para a análise multivariada de dados, especialmente quando se deseja estimar concentrações de analitos em matrizes complexas. O método foi originalmente desenvolvido por Herman Wold na década de 1970 para aplicações em economia e, posteriormente, adaptado para a quimiometria por seu filho Svante Wold, com o objetivo de lidar com variáveis altamente correlacionadas e substituir o método clássico dos mínimos quadrados (Ferreira, 2015).

Ao contrário da Regressão por Componentes Principais (*Principal Component Regression – PCR*), em que as variáveis latentes são extraídas exclusivamente com base na variância dos preditores $\mathbf{X}$, o PLS busca encontrar variáveis latentes que maximizem simultaneamente a variância em $\mathbf{X}$ e a covariância entre $\mathbf{X}$ e a variável resposta $\mathbf{y}$. Isso faz com que o modelo seja direcionado para a predição da propriedade de interesse, tornando o PLS mais eficiente e específico para cada analito (Mehmood *et al.*, 2012; Olivieri, 2017, Ferreira, 2015).

O modelo PLS representa as variáveis observadas $\mathbf{X}$ e a variável resposta $\mathbf{y}$ em termos de variáveis latentes (também chamadas de componentes ou fatores) que resumem a informação relevante. As variáveis latentes são obtidas por meio do algoritmo NIPALS, que calcula iterativamente os vetores de pesos $\mathbf{w}$, escores $\mathbf{t}$, cargas $\mathbf{p}$ e coeficientes $\mathbf{q}$, buscando maximizar a covariância entre os escores $\mathbf{t}$ e a variável resposta $\mathbf{y}$ (Ferreira, 2015).

O algoritmo PLS é desenvolvido em duas fases principais: a fase de calibração e a fase de validação.

3.4.1.1 Fase de Calibração

De acordo com Olivieri (2018), na fase de calibração do método PLS, é construído um modelo matemático inverso que relaciona as concentrações do analito com os escores das amostras, pode-se dizer, então, que se propõe estabelecer uma relação entre as variáveis espectrais e as concentrações do analito, semelhante ao utilizado na PCR, porém considerando-se tanto dados espectrais quanto informações acerca da concentração do analito.

O modelo de calibração PLS pode ser descrito da seguinte maneira:

$$\mathbf{y}_n = \mathbf{T}_A \mathbf{v}_n + \mathbf{e}$$

onde:

- $\mathbf{y}_n$ é o vetor das concentrações do analito.
- $\mathbf{T}_A$ é a matriz truncada de escores PLS, construída a partir dos dados espectrais $\mathbf{X}$ e ajustada pelas cargas $\mathbf{P}$ e pelos pesos $\mathbf{W}$.
- $\mathbf{v}_n$ é o vetor de coeficientes de regressão no espaço latente.
- $\mathbf{e}$ representa os erros residuais de modelagem.

A estimativa de $\mathbf{v}_n$ é dada por

$$\mathbf{v}_n = \mathbf{T}_A^+ \mathbf{y}_n$$

onde $\mathbf{T}_A^+$ é a pseudoinversa da matriz $\mathbf{T}_A$.

Ao contrário da PCR, em que a matriz de escores depende apenas de $\mathbf{X}$, no PLS, a matriz de escores $\mathbf{T}_A$ é calculada considerando tanto os dados espectrais quanto as concentrações do analito. A expressão específica para $\mathbf{T}_A$ é dada por:

$$\mathbf{T}_A = \mathbf{X}^T \mathbf{W}_A (\mathbf{P}_A^T \mathbf{W}_A)^{-1}$$

Nesse caso:

$\mathbf{W}_A$ são os pesos.

$\mathbf{P}_A$ são as cargas truncadas para as primeiras A colunas, responsáveis por reter a maior parte da variância espectral e de concentração.

A ortogonalidade das colunas de $\mathbf{T}_A$ assegura que os problemas de correlação presentes em $\mathbf{X}$ sejam eliminados, resultando em um modelo robusto e interpretável para análises químicas complexas.

3.4.1.2 Fase de Validação

Ainda de acordo com Olivieri (2018), na fase de validação, o modelo PLS é aplicado a novas amostras, de tal modo que os coeficientes de regressão obtidos durante a etapa de calibração são utilizados para estimar a concentração do analito, essas são ditas amostras de teste. Antes de realizar a predição, é necessário calcular os escores da amostra de teste, que resumem as informações do espectro da amostra em um espaço matemático simplificado. Esses

escores são determinados utilizando as matrizes de cargas truncadas $\mathbf{W}_A$ e $\mathbf{P}_A$, derivadas do modelo de calibração, para isso utiliza-se a seguinte formula:

$$\mathbf{t}_A = (\mathbf{W}_A^T \mathbf{P}_A)^{-1} \mathbf{W}_A^T \mathbf{x}$$

onde $\mathbf{t}_A$ representa os escores da amostra de teste, $\mathbf{W}_A$ e $\mathbf{P}_A$ são as matrizes de pesos e cargas, respectivamente, e $\mathbf{x}$ é o espectro da amostra.

Após determinar $\mathbf{t}_A$, a predição da concentração do analito (y) é realizada utilizando a seguinte equação:

$$y = \mathbf{v}_n^T \mathbf{t}_A$$

onde y é a concentração prevista do analito na amostra de teste, cujo espectro é representado pelo vetor $\mathbf{x}$ na Equação anterior.

Assim, o modelo integra informação espectral e de concentração para realizar predições confiáveis (Ferreira, 2015).

3.4.1.3 Determinação do Número Ótimo de Variáveis Latentes

Apesar das vantagens do PLS, sua aplicação requer alguns cuidados. A escolha adequada do número de variáveis latentes é crucial para garantir a precisão e a robustez do modelo. O uso excessivo de variáveis latentes pode levar ao sobreajuste (*overfitting*), onde o modelo se adapta excessivamente aos dados de calibração, capturando também o ruído e variações irrelevantes, em vez de focar nas tendências significativas. Por outro lado, um número insuficiente pode resultar em um modelo subajustado (*underfitting*), incapaz de capturar a complexidade dos dados e comprometendo sua predição (Olivieri, 2017; Chen *et al.*, 2018).

Dessa forma um aspecto importante a ser considerado no modelo PLS, é a escolha do número adequado de variáveis latentes. Esse número é determinado geralmente por validação cruzada, que pode ser realizada por técnicas como *leave-one-out* ou Monte Carlo. O número ideal de componentes é aquele que minimiza o erro de validação, como a raiz do erro quadrático médio de validação (RMSEV) ou de validação cruzada (RMSECV) (Olivieri, 2018).

Embora o PLS seja eficaz para lidar com alta dimensionalidade e multicolinearidade, ele processa o espectro como um bloco completo, o que pode diluir informações localizadas em faixas específicas. Para superar essa limitação, técnicas como o *Interval Partial Least Squares* (iPLS) e o Algoritmo das Projeções Sucessivas (SPA), combinado com Regressão Linear Múltipla (*Multiple Linear Regression* – MLR), têm sido desenvolvidas. Essas abordagens realizam a seleção de variáveis espectrais relevantes, melhorando a precisão e a interpretabilidade dos modelos em aplicações analíticas complexas (Pereira et al., 2022)

3.4.2 Regressão Linear Múltipla (*Multiple Linear Regression-MLR*)

A Regressão Linear Múltipla (MLR) é uma das técnicas de predição mais amplamente utilizadas, sendo bastante popular devido à sua simplicidade e eficácia. Ela apresenta uma ampla gama de aplicações em diversos campos, como medicina, engenharia, meio ambiente e finanças. A principal vantagem da técnica está pautada em permitir a análise da relação entre uma variável dependente e várias variáveis independentes, o que oferece maior controle sobre as análises preditivas. Esse controle torna a MLR vantajosa quando comparada a outras técnicas mais complexas, como as de aprendizado de máquina. Em termos descritivos, os coeficientes de regressão estimam o efeito das variáveis explicativas, ao controlar outras variáveis, fortalecendo a relação entre as variáveis de resposta e as explicativas (Etemadi; khasheI, 2021).

De acordo com Ferreira (2015) e com Rath, Tripathy e Tripathy (2020), a MLR considera simultaneamente diversos fatores, oferecendo uma visão mais abrangente sobre como diferentes preditores se interrelacionam, influenciando o resultado. A equação que descreve a MLR pode ser expressa como:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Onde:

Y é a variável dependente,

$X_1, X_2, \dots, X_n, \dots$, são as variáveis independentes,

β_0 é o intercepto,

$\beta_1, \beta_2, \dots, \beta_n, \dots$, são os coeficientes das variáveis,

ε é o termo de erro.

Quando se realiza a análise de dados utilizando regressão múltipla, modelos que incluem interações e termos quadráticos proporcionam uma representação mais flexível das relações entre as variáveis. Quando há interação entre variáveis, como x_1 e x_2 , o efeito de uma variável na resposta y depende do valor da outra. Esse tipo de modelo, com termos de interação, permite que as relações entre as variáveis não sejam apenas lineares, mas também mais complexas, refletindo formas variadas e realistas de interação. Esses modelos são particularmente úteis para sistemas reais, onde as relações podem ser não-lineares e dependem de múltiplos fatores (Montgomery; Peck; Vining, 2012).

A MLR também é frequentemente utilizada para prever variáveis não observadas ou para prever eventos futuros, como demonstrado em diversos estudos. Por exemplo, Rath *et al.* (2020) aplicaram a MLR para prever a evolução de casos ativos de COVID-19, enquanto Tang *et al.* (2019) usaram a técnica para avaliar o envelhecimento vascular a partir de parâmetros das ondas de pulso arterial. Outros exemplos incluem aplicações ambientais, como a previsão de consumo de energia e otimização de processos industriais (Park *et al.*, 2018; Khemet & Richman, 2018), e na análise de alimentos usou-se essa técnica para o monitoramento do frescor da carne de porco (Wei *et al.*, 2015; Wei *et al.*, 2016). Esses estudos ilustram a relevância da MLR na modelagem e previsão de fenômenos complexos, destacando seu papel fundamental no desenvolvimento de soluções preditivas eficientes e precisas.

3.4.3 Algoritmo de Projeções Sucessivas (*Successive Projections Algorithm – SPA*)

O Algoritmo de Projeções Sucessivas (SPA) é um método determinístico que se mostra eficaz na redução da multicolinearidade em dados multivariados. Ele é projetado de forma a gerar cadeias de variáveis compactas com a menor redundância possível. Dessa maneira seu objetivo principal é minimizar a multicolinearidade e eliminar variáveis redundantes, o que contribui para a criação de modelos mais simples e robustos. O algoritmo funciona de maneira sequencial, projetando variáveis originais em novos espaços ortogonais entre si. Essa abordagem intencional garante que, ao final do processo de seleção, seja obtida uma solução única e sem problemas de correlação excessiva entre as variáveis (Soares, 2010).

O SPA-MLR, é um exemplo de aplicação do SPA, sendo utilizado na regressão linear múltipla (MLR), ajudando a determinar a importância relativa das variáveis preditoras em relação à variável de interesse, tornando o modelo mais eficiente e preciso. O SPA pode ainda

ser combinado com técnicas como iSPA-PLS e iSPA-PLS-DA, que integram seleção de intervalos espectrais e corroboram com a redução de ruído do PLS, aprimorando a calibração multivariada e a classificação em modelos analíticos complexos, respectivamente. Esses algoritmos demonstram a eficácia do SPA em otimizar o desempenho preditivo, especialmente em cenários com matrizes de dados mais desafiadoras (Gomes *et al.*, 2022).

O SPA-MLR é um algoritmo estruturado em três fases, projetado para selecionar variáveis relevantes e minimizar a multicolinearidade em modelos preditivos. Na primeira fase, o algoritmo utiliza projeções sucessivas, iniciando com uma variável inicial x_j e adicionando novas variáveis a cada interação. Nesse ponto, o objetivo de maximizar os subconjuntos utilizados para cálculos de matriz inversa, sem considerar, neste momento, a relação com a variável dependente (y). Na segunda fase, é realizado um ciclo interno que seleciona o melhor subconjunto de variáveis identificado na etapa anterior, com base na minimização de uma função de custo, como o RMSEV ou RMSECV. Na terceira e última fase, os coeficientes de regressão servem como critério para eliminar variáveis restantes no subconjunto selecionado, garantindo um modelo mais eficiente e preciso (Canova *et al.*, 2023). Aplicações práticas do SPA-MLR podem ser encontrados na literatura e incluem as mais variadas análises, tais como a análise de estabilidade oxidativa, número de acidez e teor de água em biodiesel (de Lira *et al.*, 2010), a determinação de metais como Cu, Pb, Cd, Ni, Co e Zn em biocombustíveis de bioetanol (Nascimento *et al.*, 2014) e a avaliação do teor de gordura em grãos de amendoim (Sun *et al.*, 2020). Esses exemplos corroboram com a importância da técnica e demonstram sua versatilidade na otimização de modelos em diferentes contextos analíticos.

Pode-se dizer que o algoritmo iSPA-PLS é uma técnica quimiométrica avançada que combina o método de seleção de variáveis do SPA com o PLS, tornando-se assim uma ferramenta importante e amplamente utilizada para modelagem preditiva em conjuntos de dados espectroscópicos. Ele opera em duas fases principais. Na primeira fase, as variáveis da matriz $\mathbf{X}_{\text{cal}}$ são particionadas em intervalos previamente definidos, e a variável de maior norma dentro de cada intervalo é selecionada como representativa, formando uma nova matriz $\mathbf{W}_{\text{cal}}$, que possui menor dimensionalidade. Essa abordagem facilita a seleção de intervalos relevantes, cujos índices são armazenados na matriz $\mathbf{SEL}$. Na fase seguinte, diferentes combinações de intervalos são utilizadas para construir modelos PLS. Essas combinações previamente definidas na matriz $\mathbf{SEL}$, abrangem m intervalos, começando no k -ésimo intervalo, com k variando de 1 a w e m variando de 1 a $(w-1)$. Isso permite o ajuste do número de variáveis latentes,

selecionadas de forma a minimizar métricas de erro, como RMSEV ou RMSECV (Gomes *et al.*, 2013). A técnica contribui de forma eficaz para reduzir a complexidade computacional, selecionar variáveis relevantes e construir modelos robustos, sendo aplicada em trabalhos já consolidados na literatura, como em estudos para a determinação de polifenóis e umidade em chá (Diniz *et al.*, 2015), teor de antocianinas em jabuticabas (Mariani *et al.*, 2015) e gordura em hambúrgueres de frango (Krepper *et al.*, 2018), destacando-se como uma ferramenta versátil e confiável em análises químicas e espectroscópicas.

4 METODOLOGIA

4.1 Amostras

Este estudo foi composto por 41 amostras de leite de cabra puro (isto é, não adulterado), que foram preparadas aleatoriamente a partir da combinação de seis lotes provenientes de cinco rebanhos distintos de cabras das raças Alpina Americana, Anglo-Nubiana, Moxotó, Murciana e Parda. Adicionalmente, 95 amostras adulteradas foram preparadas pela adição de leite de vaca de quatro lotes das raças Sindi, Guzerá e Girolando, em concentrações variadas (5, 7, 10, 20, 25, 30, 40, 50 e 75 g/100 g).

As amostras de leite de cabra e de vaca foram obtidas junto a produtores das regiões do Agreste, Cariri e Zona da Mata, no estado da Paraíba, Brasil, sendo os animais caracterizados por variações de idade, estágio de lactação e sistema de criação, incluindo manejo intensivo, semi-intensivo e extensivo.

Os lotes individuais de leite de cabra e de vaca foram coletados por ordenha manual supervisionada, a fim de garantir a integridade das amostras (isto é, ausência de adulteração) e evitar contaminações externas. As amostras de leite foram posteriormente submetidas a tratamento térmico a 65 °C por 30 minutos e armazenadas a 4 °C até a análise. Processos térmicos acima de 65 °C têm mostrado promover interações entre proteínas do soro desnaturadas e proteínas da membrana do glóbulo de gordura do leite (MFGM), além de facilitar a dissociação de proteínas nativas associadas à MFGM (Wang et al., 2024; Wiking et al., 2022). A aplicação de tratamento térmico a 65 °C demonstrou reduzir significativamente a viabilidade de bactérias psicrótroficas e eliminar coliformes, melhorando assim a qualidade microbiológica e aumentando a vida útil do leite armazenado a 4 ± 2 °C após a termização (Wang et al., 2024; Wiking et al., 2022; Molska et al., 1989).

As amostras adulteradas, compostas por alíquotas de 10 g de misturas de leite de cabra e vaca, foram preparadas em duplicata utilizando as amostras de leite puro termicamente tratadas, antes do resfriamento. A homogeneização das amostras foi realizada por inversões manuais repetidas por três vezes, em tubos Falcon de 50 mL, seguidas de armazenamento a 4 °C. O teor de gordura das amostras foi determinado pelo método de Gerber, conforme o Método Oficial 2000.18 da AOAC (Pereira et al., 2021), enquanto as medições no infravermelho próximo (NIR) foram realizadas em triplicata, conforme descrito a seguir.

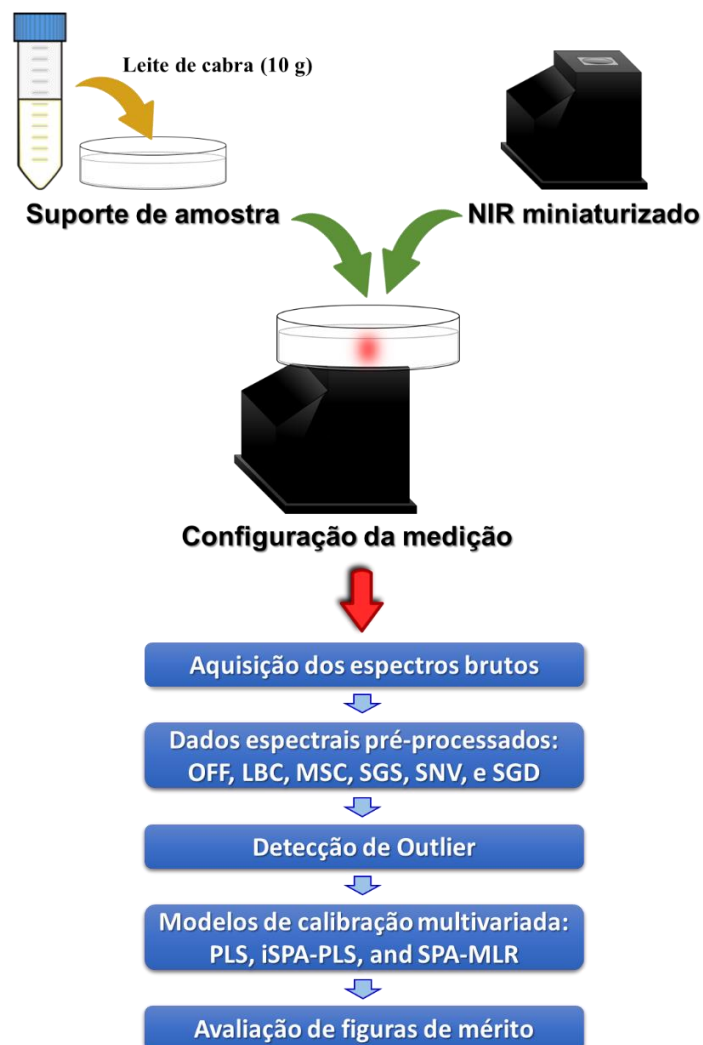
4.2 Medições por NIR

Antes das medições por espectroscopia NIR, as amostras foram deixadas em repouso

até atingirem a temperatura ambiente ($23 \pm 1^\circ\text{C}$) para garantir estabilização térmica. Em seguida, os tubos Falcon de armazenamento foram invertidos manualmente várias vezes para evitar a separação de fases. Posteriormente, 10 gramas de cada amostra foram transferidos para placas de Petri de poliestireno ($60\text{ mm} \times 15\text{ mm}$), utilizadas como suportes de amostra. Essas placas foram então posicionadas diretamente sobre a janela óptica do espectrômetro para aquisição dos espectros de reflectância.

Uma placa de Petri contendo pó de politetrafluoretileno (PTFE) foi usada como referência de reflectância (branco) durante a aquisição espectral, para evitar interferências espectrais causadas pelo suporte da amostra. A montagem experimental está representada na Figura 1.

Figura 1 – Esquema do procedimento analítico para aquisição e tratamento de espectros NIR de amostras de leite



Vale destacar que este protocolo padronizado evitou de forma eficaz a separação de fases, o deslocamento de bandas espectrais e a saturação de picos, garantindo assim a reprodutibilidade e confiabilidade das medições espectroscópicas, conforme previamente relatado (Pereira et al., 2020, 2021).

Espectros NIR na faixa de 900–1650 nm foram adquiridos utilizando um instrumento portátil de NIR (DLP NIRscan Nano, Texas Instruments, EUA). O espectrômetro portátil utilizado neste estudo integra um dispositivo microespelhado digital (DMD), fabricado com tecnologia de sistemas microeletromecânicos (MEMS), o que melhora a sensibilidade e a relação sinal-ruído, apesar da menor resolução espectral (900–1650 nm) em comparação com instrumentos de bancada (900–2500 nm). Sua portabilidade, custo acessível e facilidade de uso favorecem o controle de qualidade *in loco*, beneficiando especialmente pequenos produtores com acesso limitado à infraestrutura laboratorial (Grabska, Beć, & Huck, 2024; Giussani, Gorla, & Riu, 2022).

A aquisição espectral foi realizada com resolução nominal de 3,5 nm, resultando em dados comprimidos em 210 pontos de comprimento de onda. Cada espectro foi obtido pela integração de 32 varreduras utilizando a transformada de Hadamard, sob temperatura ambiente (23 ± 1 °C). Para a modelagem quimiométrica subsequente, apenas os espectros médios de cada amostra foram considerados.

4.3 Procedimento quimiométrico

Uma representação esquemática das etapas envolvidas na análise quimiométrica foi apresentada anteriormente na Figura 1. Os espectros NIR foram pré-processados utilizando diferentes técnicas: (a) correção de offset (OFF), (b) correção linear de linha de base (LBC), (c) correção de espalhamento multiplicativo (MSC), (d) suavização de Savitzky-Golay com janela de 11 pontos (SGS), (e) variação normal padrão (SNV) e (f) primeira derivada de Savitzky-Golay com polinômio de 2ª ordem e janela de 11 pontos (SGD).

Antes da construção dos modelos de calibração multivariada, foi realizada uma investigação de outliers utilizando o gráfico de influência, definido como $3*k/n$ (onde k é o número de fatores e n é o número de amostras) versus resíduos studentizados (com limites a 95% de nível de confiança). Em seguida, o conjunto de dados foi dividido em conjuntos de calibração (75%) e predição (25%) utilizando o algoritmo de Kennard-Stone. Modelos de regressão PLS, iSPA-PLS e SPA-MLR foram construídos e validados utilizando validação

cruzada completa para otimizar o número de variáveis latentes (LV) a serem incluídas no modelo de calibração. O desempenho dos modelos foi avaliado em termos do coeficiente de determinação (R^2_{CV}) e da raiz do erro quadrático médio da validação cruzada (RMSECV) do conjunto de calibração. Além disso, a capacidade preditiva dos modelos também foi avaliada por meio da raiz do erro quadrático médio de predição (RMSEP), erro relativo de predição (REP) e da razão de desempenho do desvio (RPD) (Gomes et al., 2022; Esbensen, Geladi, & Larsen, 2014). Por fim, o teste da região elíptica de confiança conjunta (EJCR) foi utilizado para investigar a presença ou ausência de viés, ou seja, para avaliar a exatidão dos modelos no conjunto de predição. Nesse teste, realiza-se inicialmente um ajuste por mínimos quadrados ordinários (OLS); em seguida, os valores estimados do intercepto (a) e do coeficiente angular (b) são comparados com seus valores ideais de $a = 0$ e $b = 1$, respectivamente (Duarte et al., 2022).

Todos os procedimentos quimiométricos foram realizados utilizando rotinas desenvolvidas internamente no Matlab® 2019a (Mathworks, EUA).

5 RESULTADOS E DISCUSSÃO

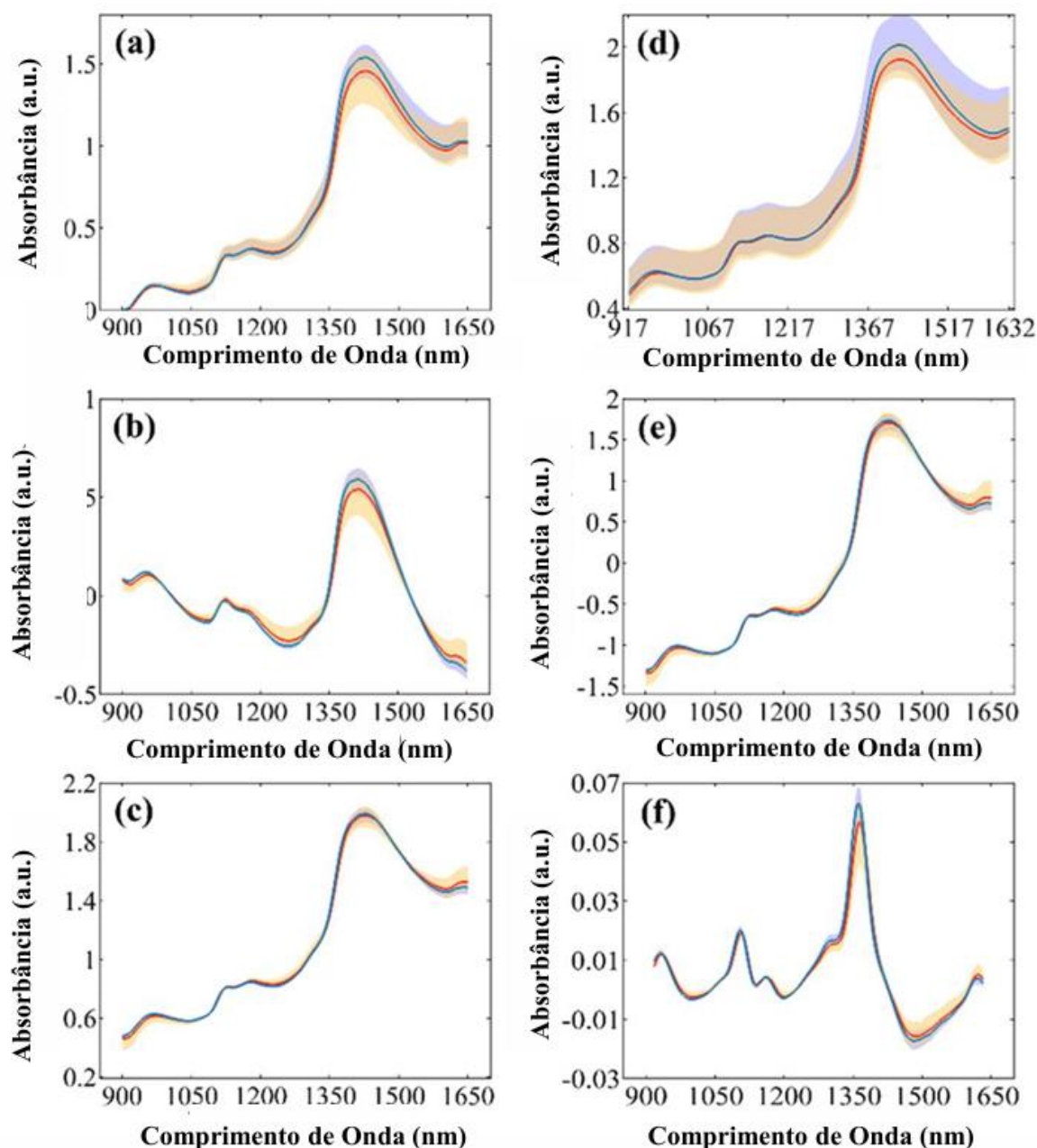
5.1 Investigação espectral

A Figura 2 mostra os espectros médios NIR pré-processados com OFF (Fig. 2a), LBC (Fig. 2b), MSC (Fig. 2c), SGS (Fig. 2d), SNV (Fig. 2e) e SGD (Fig. 2f) para as amostras de leite de cabra puro (linha sólida laranja) e adulterado (linha sólida azul). As regiões sombreadas ao redor dos espectros médios indicam a variabilidade máxima observada dentro da faixa espectral para as amostras de leite de cabra puro (sombreamento laranja) e adulterado (sombreamento azul).

A região espectral NIR estudada fornece informações valiosas sobre a composição molecular de matrizes complexas devido à absorção de sobretons específicos e bandas de combinação associadas a vibrações fundamentais. Neste estudo, a banda proeminente e ampla observada em torno de 1450 nm é atribuída ao primeiro sobretom da vibração de estiramento O–H da água. Essa característica domina os espectros devido à alta concentração de água em amostras biológicas, refletindo sua forte interação com a radiação NIR. Uma análise dos espectros médios, considerando todas as técnicas de pré-processamento aplicadas, revela que as variações espectrais mais significativas, indicadas pelas regiões sombreadas, estão predominantemente concentradas na banda em aproximadamente 1450 nm. Esse efeito é mais evidente na Figura 2d, destacando a resposta espectral distinta nessa região. Essa variabilidade proeminente, ligada principalmente às diferenças no teor de água entre as amostras de leite, provavelmente decorre da heterogeneidade na composição dos rebanhos, idade dos animais, estágios de lactação e sistemas de criação.

Bandas de absorção menores detectadas entre 1100 e 1200 nm (Fig. 2a–2f) estão associadas ao segundo sobretom das vibrações de estiramento –CH, indicativas do teor de gordura das amostras. Essas bandas fornecem informações essenciais sobre a composição lipídica e são altamente relevantes para a avaliação dos níveis de gordura em produtos lácteos. Além disso, a região espectral em torno de 1630 nm corresponde ao primeiro harmônico das vibrações de estiramento N–H e à banda de combinação amida II. Essa característica está principalmente relacionada ao conteúdo proteico da amostra, com alta variação observada nas regiões sombreadas entre 1600 e 1650 nm (Fig. 2a–2f). Essas assinaturas espectrais distintas possibilitam a análise simultânea de componentes-chave como água, gordura e proteína, destacando a robustez da espectroscopia NIR para análises composicionais em matrizes lácteas complexas (He et al., 2023; Pereira et al., 2021, 2020; Teixeira et al., 2020).

Figura 2 Espectros médios NIR pré-processados com OFF (Fig. 2a), LBC (Fig. 2b), MSC (Fig. 2c), SGS (Fig. 2d), SNV (Fig. 2e) e SGD (Fig. 2f) para as amostras de leite de cabra puro (linha sólida laranja) e adulterado (linha sólida azul). As áreas sombreadas ao redor dos espectros médios NIR representam a variabilidade máxima observada em cada canal analítico para as duas classes estudadas.



Fonte: Elaborada pelo autor, 2025.

5.2 Quantificação do teor de gordura

O teor lipídico do leite de cabra geralmente varia entre 3% e 5%, sendo composto predominantemente por triacilgliceróis, e apresenta variação significativa ao longo da lactação.

Neste estudo, algumas amostras apresentaram teor de gordura anormalmente baixo (<2 g/100 g), provavelmente influenciado por múltiplos fatores. As amostras foram coletadas de diferentes produtores na Paraíba, englobando condições ambientais diversas, idades dos animais, estágios de lactação e sistemas de alimentação variados. Nutricionalmente, dietas ricas em carboidratos fermentáveis ou com baixo teor de fibra podem induzir à depressão do teor de gordura do leite (MFD), enquanto ácidos graxos insaturados podem interferir na bioidrogenação ruminal, inibindo a síntese de gordura. Além disso, flutuações naturais no teor de gordura ocorrem devido ao estágio de lactação, estresse térmico, disponibilidade de água e práticas de manejo. Por exemplo, a ordenha incompleta pode resultar em níveis mais baixos de gordura nas frações iniciais do leite. Esses fatores explicam a variabilidade observada, enfatizando a influência multifatorial das condições genéticas, nutricionais e ambientais na composição lipídica. Consequentemente, a distribuição heterogênea do teor de gordura reflete diretamente esses determinantes, sustentando a faixa analítica estabelecida e destacando a variabilidade entre populações caprinas e sistemas de manejo (Leskinen et al., 2019; Rico & Harvatine, 2013).

A Tabela 4 apresenta os resultados obtidos para a quantificação do teor de gordura nas amostras de leite de cabra e suas misturas com leite de vaca. Como pode ser observado, todos os modelos multivariados apresentaram bom ajuste na etapa de calibração, com coeficientes de correlação (r_{cv}) superiores a 0,92 e valores de RMSECV inferiores a 0,26 g/100 g. No entanto, na etapa de predição, os modelos PLS, iSPA-PLS e SPA-MLR construídos com SGS apresentaram a melhor capacidade preditiva, com coeficientes de correlação (r_{pred}) entre 0,97 e 0,98 e valores de RMSEP entre 0,12 e 0,13 g/100 g. Além disso, os valores de RPD foram superiores a 4,09 e os de REP inferiores a 2,86%, sem resultados enviesados (isto é, os valores de t_{calc} foram inferiores ao t_{crit} ($=1,71$)).

Para ilustrar os melhores resultados da quantificação do teor de gordura obtidos pelos diferentes algoritmos de calibração multivariada usando os espectros NIR pré-processados com suavização de Savitzky-Golay, foi realizada uma etapa de detecção de outliers, conforme mostrado nas Figuras 3a–3c. Como se pode observar, não foram detectados outliers acima dos valores críticos de influência e resíduos studentizados simultaneamente. Em seguida, a etapa de otimização dos modelos envolveu a seleção do número ideal de variáveis latentes no modelo SGS/PLS, dos intervalos no SGS/iSPA-PLS e das variáveis individuais no SGS/SPA-MLR, conforme apresentado nas Figuras 3d–3f. Embora o número de variáveis latentes no SGS/PLS possa parecer excessivamente alto, vale ressaltar a significativa complexidade dessa matriz alimentar estudada, incluindo diferentes raças de cabras, sistemas de criação, idades dos

animais, estágios de lactação, condições edafoclimáticas e níveis variados de mistura de leite (leite de cabra + leite de vaca). Sabe-se também que a construção interna de modelos geralmente requer de 2 a 3 fatores e de 1 a 2 fatores para integrar cada nova fonte de variação analítica (Pomerantsev & Rodionova, 2018). Apesar de parecer um resultado sobreajustado para o PLS com 14 variáveis latentes, destaca-se que o algoritmo PLS empregado neste estudo incorpora o teste de Haaland-Thomas, que remove sistematicamente as variáveis latentes avaliando se há melhorias estatisticamente significativas com variáveis adicionais. A Figura 3d mostra que de 15 a 17 variáveis latentes apresentaram valores de RMSECV mais baixos, mas sem diferença estatística significativa em relação a 14 variáveis latentes. Em geral, o sobreajuste resultaria em pior desempenho preditivo em comparação com calibração e validação. No entanto, nossos resultados confirmam que as 14 variáveis latentes selecionadas são justificáveis, dada a complexidade do conjunto de dados.

Para otimizar ainda mais o modelo SGS/PLS, utilizou-se a estratégia de importância das variáveis na projeção (*Variable Importance in Projection* – VIP), fornecendo uma avaliação abrangente da relevância de cada variável ao longo do espectro. Conforme mostrado na Figura 4, todas as variáveis apresentaram escores VIP superiores a 1, indicando alta significância para o modelo. Além disso, observa-se um aumento gradual nos escores VIP ao longo do espectro, sugerindo que regiões espectrais específicas, como aquelas selecionadas pelo modelo SGS/iSPA-PLS (Fig. 4f), possuem maior importância na modelagem dos dados. Estes resultados também são corroborados pelo modelo SGS/SPA-MLR, no qual as variáveis selecionadas (1125, 1100, 935, 1340, 917, 1254, 1391, 1366 e 1420 nm) (Fig. 3e) coincidem parcialmente com as regiões de maiores escores VIP, reforçando sua relevância no modelo PLS.

Adicionalmente, os gráficos de valores previstos versus valores de referência para as amostras de calibração (círculos verdes) e predição (quadrados cianos) ilustram o ajuste dos modelos PLS, iSPA-PLS e SPA-MLR usando espectros NIR pré-processados com SGS nas Figuras 3g–3i, respectivamente. Como se observa, embora as amostras pareçam estar bem distribuídas aleatoriamente em ambos os lados das linhas bisetrizes, há um desvio das elipses de confiança conjuntas em relação ao ponto teórico ideal (1,0) nos modelos SGS/PLS (Fig. 4j) e SGS/SPA-MLR (Fig. 3k). Esse desvio pode ser atribuído às características espectrais dos modelos selecionados e às propriedades inerentes da espectroscopia NIR. O modelo SGS/PLS incorporou bandas de absorção relevantes para lipídios, mas também incluiu regiões espectrais com correlações fracas ou inespecíficas com o teor de gordura. A inclusão dessas regiões adicionais provavelmente introduziu regiões não-informativas e/ou informações redundantes, reduzindo a especificidade do modelo e contribuindo para o desvio em relação ao ponto ideal.

Esses resultados sugerem que, embora o PLS capte padrões espectrais amplos, pode carecer da precisão necessária para a quantificação exata da gordura sem uma estratégia apropriada de seleção de variáveis.

Tabela 4. Resultados obtidos para a quantificação do teor de gordura em amostras de leite de cabra e suas misturas com leite de vaca utilizando PLS, iSPA-PLS e SPA-MLR.

Parâmetros									
Modelos	Técnica de pre-processamento	RMSECV (g/100g)	r_{cv}	RMSEP (g/100g)	r_{pred}	RPD	REP (%)	t_{cal} ($t_{crit} = 1.71$)	LV or NV
PLS		0,24	0,94	0,20	0,93	2,58	4,54	1,27	13
iSPA-PLS (20/16)	OFF	0,18	0,96	0,18	0,94	2,94	3,99	1,16	13
SPA-MLR		0,20	0,96	0,22	0,57	0,97	4,66	1,72	17
PLS		0,26	0,92	0,22	0,92	2,44	4,84	1,05	12
iSPA-PLS (15/8)	LBC	0,20	0,95	0,18	0,95	3,02	3,91	0,50	12
SPA-MLR		0,19	0,96	0,33	0,80	1,62	7,28	1,13	12
PLS		0,26	0,93	0,25	0,89	2,10	5,54	0,09	10
iSPA-PLS (10/5)	MSC	0,18	0,96	0,26	0,91	2,05	5,69	0,77	9
SPA-MLR		0,23	0,94	0,28	0,87	1,91	6,10	0,90	16
PLS		0,18	0,96	0,12	0,98	4,30	2,72	0,54	14
iSPA-PLS (10/3)	SGS	0,14	0,98	0,12	0,97	4,32	2,70	0,07	9
SPA-MLR		0,14	0,98	0,13	0,97	4,09	2,86	0,02	10
PLS		0,17	0,97	0,12	0,97	4,30	2,71	0,56	7
iSPA-PLS (15/12)	SNV	0,16	0,97	0,14	0,97	3,90	2,98	0,67	7
SPA-MLR		0,16	0,97	0,16	0,95	3,22	3,62	0,07	17
PLS		0,23	0,94	0,20	0,93	2,63	4,50	0,19	11
iSPA-PLS (15/7)	SGD	0,17	0,97	0,15	0,96	3,55	3,33	1,86	11
SPA-MLR		0,19	0,96	0,25	0,89	2,16	5,47	0,51	15

PLS: Mínimos Quadrados Parciais; SPA-MLR: Algoritmo de Projeções Sucessivas para seleção de variáveis em Regressão Linear Múltipla; iSPA-PLS: Algoritmo de Projeções Sucessivas para seleção de intervalos em Mínimos Quadrados Parciais; RMSECV: raiz do erro quadrático médio da validação cruzada; r_{cv} : coeficiente de correlação da validação cruzada; RMSEP: raiz do erro quadrático médio da predição; r_{pred} : coeficiente de correlação da predição; RPD: razão do desempenho do desvio; REP: erro relativo de predição; LV: variáveis latentes utilizadas no modelo; NV: número de variáveis individuais selecionadas. O número de intervalos selecionados no iSPA-PLS é indicado entre parênteses (por exemplo, 20/16 significa 16 intervalos selecionados pelo algoritmo SPA de um total de 20).

Fonte: Elaborada pelo autor, 2025.

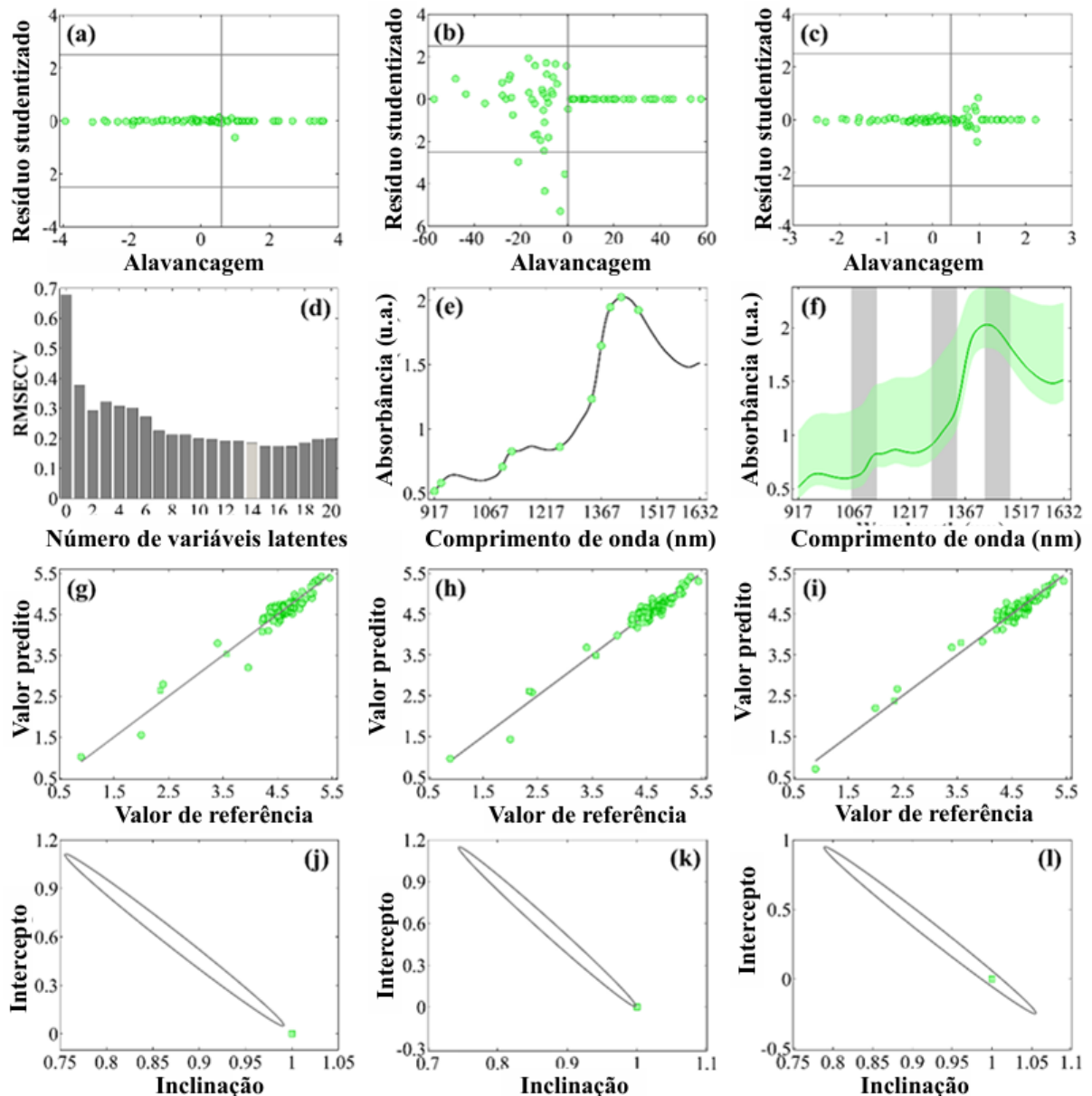
De modo semelhante, o modelo SGS/SPA-MLR, apesar de selecionar comprimentos de onda representativos, apresentou predição enviesada (Fig. 3k). Esse viés pode ser atribuído às características intrínsecas dos espectros NIR, cujas bandas de absorção amplas e sobrepostas dificultam o isolamento de contribuições moleculares específicas. Ademais, a menor resolução espectral dos instrumentos NIR portáteis em comparação com sistemas de bancada pode ter aumentado a colinearidade espectral, reduzindo a eficácia da seleção de variáveis e resultando em uma elipse de confiança que não abrange o ponto ideal.

Por outro lado, o modelo SGS/iSPA-PLS demonstrou desempenho preditivo superior ao selecionar três intervalos espectrais-chave (1060–1128 nm, 1200–1400 nm e 1420–1488 nm) fortemente associados à gordura e a outros componentes do leite. A faixa de 1060–1128 nm corresponde ao segundo sobretom das vibrações de estiramento O–H e N–H, oriundas principalmente da água e estruturas proteicas. A região de 1200–1400 nm está ligada a bandas de combinação do estiramento C–H nos grupos CH₂ e CH₃, enquanto o intervalo 1420–1488 nm corresponde aos segundos sobretons das vibrações de estiramento C–H em grupos metila (–CH₃) e metileno (–CH₂–), característicos dos lipídios. A seleção direcionada dessas regiões quimicamente relevantes eliminou regiões não-informativas e aumentou a robustez do modelo, resultando em uma elipse de confiança conjunta que abrange o ponto ideal (1,0) (Fig. 4l). Esse resultado indica a ausência de erro sistemático (isto é, ausência de viés significativo), conclusão corroborada por um teste *t* com nível de significância $\alpha = 0,05$ (Tab. 3).

Comparativamente, o modelo SGS/iSPA-PLS empregou 9 variáveis latentes para construir o modelo e alcançar o melhor desempenho preditivo, obtendo r_{pred} de 0,97, RMSEP de 0,12 g/100 g, RPD de 4,32 e REP de apenas 2,70%, além de apresentar a menor elipse de confiança (Fig. 3l).

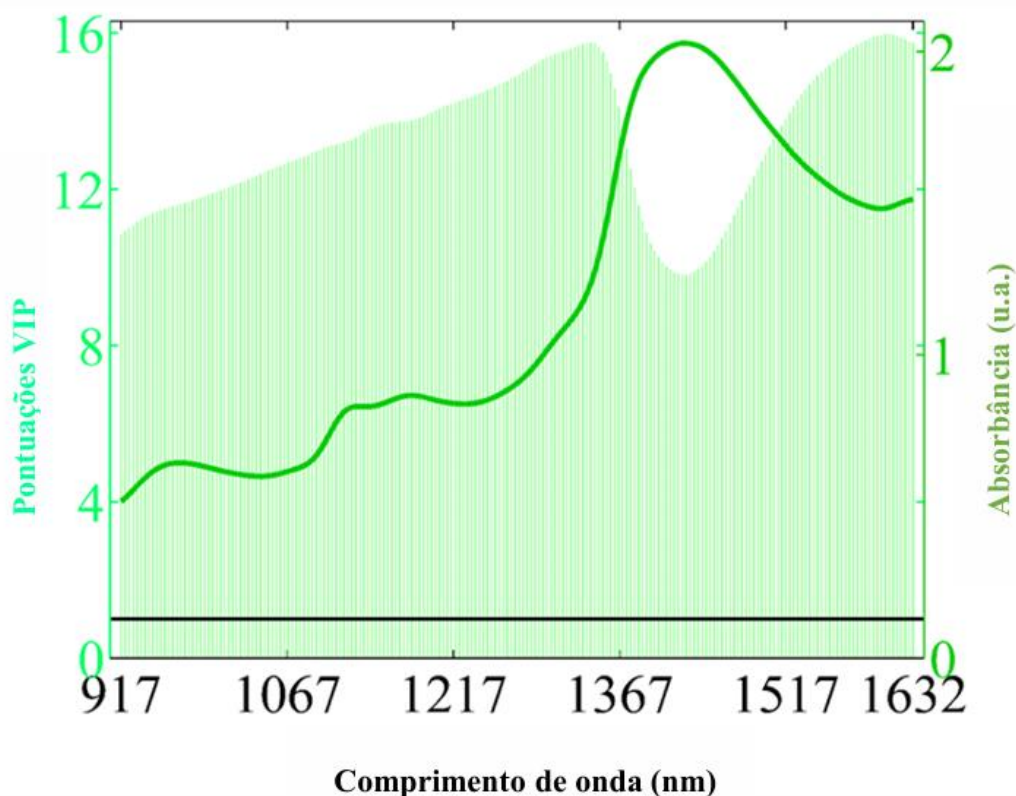
Em relação à literatura, Pereira et al. (2020) utilizaram um espectrofotômetro NIR de bancada (operando entre 1000–2500 nm) e obtiveram r_{pred} de 0,96, RMSEP de 0,20 g/100 g, RPD de 3,36 e REP de 6,37%, utilizando 11 variáveis latentes. Apesar do desempenho inferior, eles aplicaram iSPA-PLS aos espectros NIR pré-processados com média móvel de 13 pontos e correção de linha de base linear, selecionando 9 dos 20 intervalos na faixa de 1250–2000 nm. Esses resultados confirmam que as informações espectrais mais relevantes para a quantificação do teor de gordura podem ser capturadas com sucesso utilizando um espectrofotômetro NIR portátil. Além disso, destacam a eficácia da seleção de variáveis baseada em intervalos para aumentar a precisão e confiabilidade dos resultados de quantificação.

Figura 3 Resultados relacionados à determinação do teor de gordura em amostras de leite de cabra e suas misturas com leite de vaca utilizando os espectros NIR pré-processados com suavização Savitzky-Golay e janela de 11 pontos (SGS): (a-c) Gráficos de influência, que ilustram as distâncias entre os resíduos e o centro dos resultados para a construção dos modelos PLS, SPA-MLR e iSPA-PLS, respectivamente. (d-f) Etapa de otimização do número de variáveis latentes no PLS, das variáveis individuais no SPA-MLR, e dos intervalos no iSPA-PLS, respectivamente. (g-i) Gráficos de valores preditos versus valores de referência, que incluem as amostras de calibração (círculos verdes) e as amostras de predição (quadrados verdes).



Fonte: Elaborada pelo autor, 2025.

Figura 4. Valores de escores VIP (linhas verticais verdes) para cada variável incluída no modelo PLS para a determinação do teor de gordura em amostras de leite de cabra. Valores acima de 1 (linha preta contínua) indicam as variáveis importantes incluídas no modelo.



Fonte: Elaborada pelo autor, 2025.

5.3 Quantificação do teor de adulteração com leite de vaca

A Tabela 5 apresenta os resultados obtidos para a quantificação de leite de vaca em leite de cabra. A suavização Savitzky-Golay foi novamente a melhor técnica de pré-processamento para alcançar o melhor desempenho. Todos os modelos utilizando SGS apresentaram os menores valores de RMSE e os maiores coeficientes de correlação, tanto na etapa de calibração quanto na de predição. Em outras palavras, os valores de RMSE foram inferiores ao menor nível de adulteração de 5 g/100 g, e os coeficientes de correlação foram superiores a 0,98. Além disso, os valores de RPD foram muito altos ($> 7,9$), os valores de REP ficaram em torno de 8%, e não apresentaram viés segundo o teste t ($\alpha = 0,05$). Vale destacar que os modelos baseados em SPA melhoraram os resultados do PLS, especialmente pela redução dos valores de REP.

Tabela 5. Resultados obtidos para a quantificação de leite de vaca em leite de cabra utilizando PLS, iSPA-PLS e SPA-MLR.

Parâmetros									
Modelos	Técnicas de pré-processamento	RMSECV (g/100g)	r_{cv}	RMSEP (g/100g)	r_{pred}	RPD	REP (%)	t_{cal} ($t_{crit} = 1.71$)	LV or NV
PLS		7,96	0,92	2,37	0,99	8,12	8,88	1,40	12
iSPA-PLS (25/16)	OFF	6,76	0,95	2,64	0,99	7,28	9,90	0,25	12
SPA-MLR		8,25	0,92	5,69	0,96	3,38	21,34	1,62	8
PLS		8,24	0,92	2,78	0,99	6,94	10,39	0,57	11
iSPA-PLS (5/3)	LBC	7,84	0,93	2,79	0,99	6,90	10,44	0,22	11
SPA-MLR		8,27	0,92	5,04	0,97	3,82	18,86	1,11	13
PLS		8,20	0,92	2,92	0,99	6,58	10,95	0,94	11
iSPA-PLS (20/9)	MSC	6,62	0,95	3,38	0,98	5,70	12,66	0,56	11
SPA-MLR		9,78	0,88	7,21	0,93	2,67	26,99	1,14	8
PLS		4,29	0,98	2,70	0,99	7,92	8,58	0,71	6
iSPA-PLS (20/16)	SGS	4,15	0,98	2,15	0,99	8,96	8,04	0,67	6
SPA-MLR		3,85	0,98	2,22	0,99	8,69	8,30	1,46	9
PLS		8,54	0,91	2,87	0,99	6,70	10,75	1,19	11
iSPA-PLS (10/9)	SNV	8,39	0,92	2,84	0,99	6,78	10,63	0,87	11
SPA-MLR		9,40	0,89	7,04	0,93	2,74	26,36	1,00	9
PLS		8,16	0,92	3,71	0,98	5,20	13,87	1,70	9
iSPA-PLS (20/14)	SGD	7,39	0,94	3,67	0,98	5,25	13,72	0,65	9
SPA-MLR		8,16	0,92	5,80	0,96	3,32	21,70	1,90	9

PLS: Mínimos Quadrados Parciais; SPA-MLR: Algoritmo de Projeções Sucessivas para seleção de variáveis em Regressão Linear Múltipla; iSPA-PLS: Algoritmo de Projeções Sucessivas para seleção de intervalos em Mínimos Quadrados Parciais; RMSECV: raiz do erro quadrático médio da validação cruzada; r_{cv} : coeficiente de correlação da validação cruzada; RMSEP: raiz do erro quadrático médio da predição; r_{pred} : coeficiente de correlação da predição; RPD: razão do desempenho do desvio; REP: erro relativo de predição; LV: variáveis latentes utilizadas no modelo; NV: número de variáveis individuais selecionadas. O número de intervalos selecionados no iSPA-PLS é indicado entre parênteses (por exemplo, 20/16 significa 16 intervalos selecionados pelo algoritmo SPA de um total de 20).

Fonte: Elaborada pelo autor, 2025.

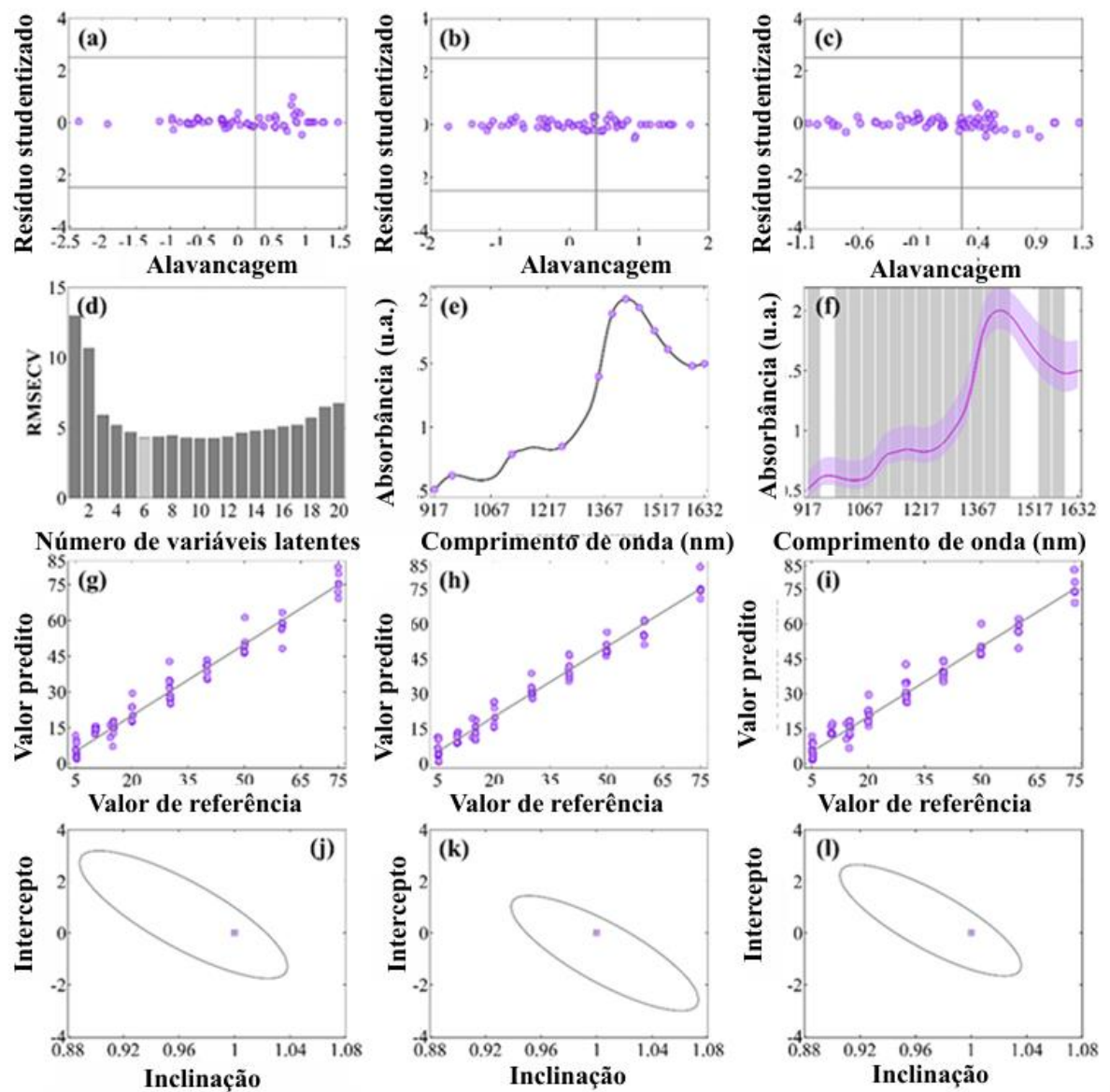
Novamente, os escores VIP foram utilizados para otimizar a seleção de variáveis, garantindo que apenas as regiões espectrais mais relevantes fossem utilizadas no modelo SGS/PLS. Conforme mostrado na Figura 6, os escores VIP apresentaram um aumento notável de 3,74 para 4,97 na faixa espectral de 917–1100 nm, uma região associada principalmente aos sobretons de estiramento O–H da água e de compostos orgânicos, incluindo proteínas e carboidratos. Na faixa de 1100–1400 nm, os escores VIP aumentaram ainda mais, de 6,0 para 6,5, provavelmente correspondendo às bandas de estiramento C–H em lipídios e grupos CH_2/CH_3 . Na região de 1400–1632 nm, os escores VIP excederam 7,0, atingindo o valor máximo de 7,37. Essa faixa espectral é comumente associada a vibrações harmônicas secundárias C–H, indicando uma forte correlação com lipídios e outros compostos orgânicos de alto peso molecular. Por outro lado, as variáveis selecionadas pelo modelo SGS/SPA-MLR (917, 1254, 1351, 1387, 1423, 1459, 1499, 1534 e 1599 nm) apresentaram forte concordância com regiões de altos escores VIP, validando ainda mais o processo de seleção. Por exemplo, o comprimento de onda de 917 nm coincide com uma região de alto valor de VIP associada à água e compostos orgânicos. Em contraste, a faixa de 1351 a 1599 nm se alinha com fortes bandas de estiramento C–H características de lipídios. Em comparação, o modelo SGS/iSPA-PLS utilizou 6 variáveis latentes para a construção do modelo e selecionou 16 dos 20 intervalos. Como esperado, a alta complexidade e similaridade entre as amostras de leite de cabra e de vaca exigem a seleção de múltiplos intervalos espectrais, particularmente aqueles associados ao teor de água, proteína e gordura, que requerem mais informações analíticas para quantificar a adulteração com precisão. No entanto, a seleção de intervalos no modelo SGS/iSPA-PLS melhora a precisão preditiva ao eliminar variáveis não-informativas, reduzindo o erro relativo de predição em mais de 0,5% em comparação ao modelo SGS/PLS.

Para ilustração, os modelos SGS/PLS, SGS/iSPA-PLS e SGS/SPA-MLR para quantificação de leite de vaca não apresentaram outliers (Figuras 5a–5c) e foram construídos após a etapa de otimização do número de variáveis latentes no PLS e variáveis individuais no MLR (Figuras 5d–5f). Todos esses modelos apresentaram bom ajuste, com os dados de calibração (círculos roxos) e de predição (quadrados bege) distribuídos aleatoriamente em ambos os lados da linha bissetriz. As Figuras 5g–5i ilustram os gráficos de valores preditos versus valores de referência, enquanto as Figuras 5h–5l mostram as elipses de confiança dos modelos de predição contendo o ponto ideal teórico para os modelos PLS, iSPA-PLS e SPA-MLR, respectivamente.

Com relação à literatura, apenas Pereira et al. (2020) relataram valores de RPD e REP

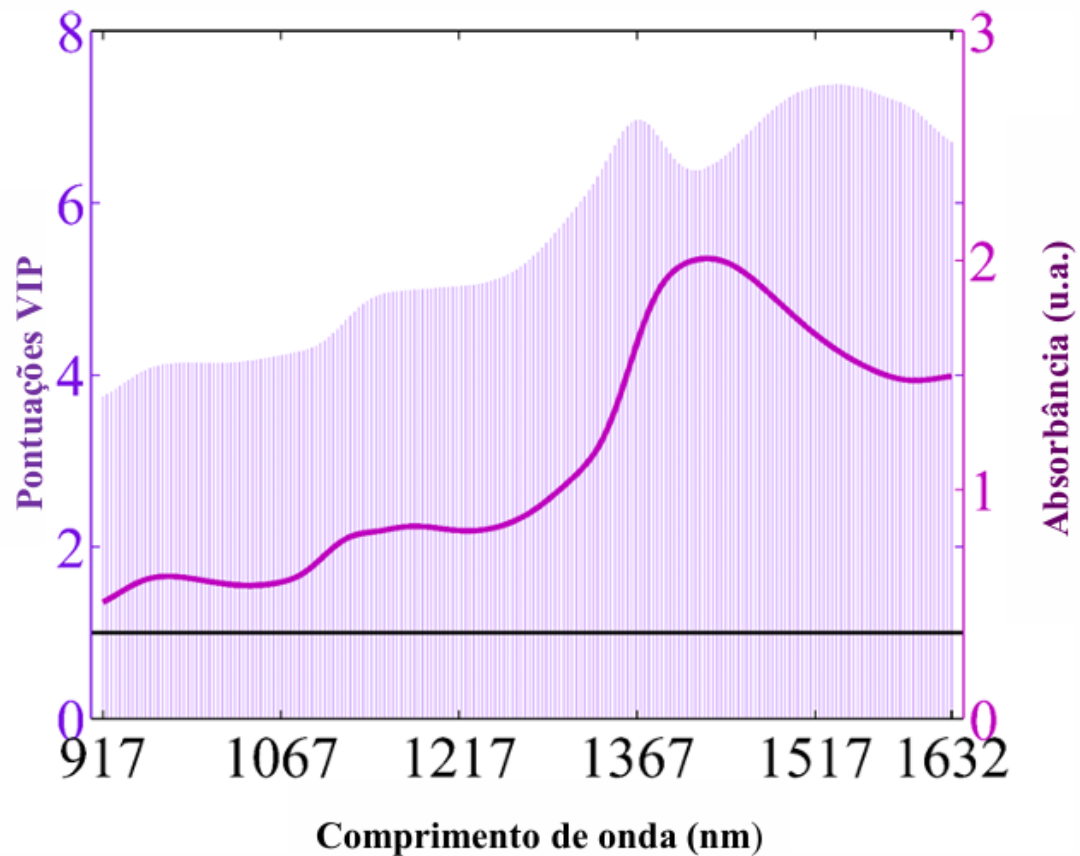
além dos valores amplamente utilizados de RMSE e coeficiente de correlação. No conjunto de predição, eles obtiveram um r_{pred} de 0,99, RMSEP de 3,66 g/100 g, RPD de 10,0 e REP de 11,2%, empregando 14 variáveis latentes e selecionando 18 dos 20 intervalos. Em contraste, o presente estudo se destaca por alcançar um r_{pred} de 0,98 e RMSEP de 4,15 g/100 g na etapa de calibração, utilizando apenas 6 variáveis latentes e selecionando 16 dos 20 intervalos (Figura 5f) para o modelo SGS/iSPA-PLS. Além disso, seu desempenho preditivo atingiu r_{pred} de 0,99, RMSEP de 2,15 g/100 g, RPD de 8,96 e REP de 8,04%. A forte concordância entre as amostras de calibração (círculos) e predição (quadrados) e a respectiva elipse de confiança contendo o ponto ideal teórico estão ilustradas nas Figuras 5i e 5l, respectivamente. Mais uma vez, os intervalos selecionados neste estudo reforçam a viabilidade do uso de um dispositivo NIR portátil como alternativa robusta e de baixo custo aos instrumentos de bancada, destacando seu potencial para aplicações em campo na avaliação da qualidade de alimentos.

Figura 5 Aspectos relacionados à determinação da adulteração com leite de vaca em amostras de leite de cabra utilizando os espectros NIR pré-processados com suavização Savitzky-Golay e janela de 11 pontos (SGS): (a-c) Gráficos de influência que mostram as distâncias dos resíduos ao centro dos resultados para a construção dos modelos PLS, SPA-MLR e iSPA-PLS, respectivamente. (d-f) Etapa de otimização do número de variáveis latentes no PLS, das variáveis individuais no SPA-MLR, e dos intervalos no iSPA PLS. (g-i) Gráficos de valores preditos versus valores de referência, mostrando a distribuição das amostras de calibração (círculos roxos) e de predição (quadrados roxos).



Fonte: Elaborada pelo autor, 2025.

Figura 6. Valores de escores VIP (linhas verticais verdes) para cada variável incluída no modelo PLS para a determinação de leite de vaca em amostras de leite de cabra. Valores acima de 1 (linha preta contínua) indicam as variáveis importantes incluídas no modelo.



Fonte: Elaborada pelo autor, 2025.

É importante observar que este estudo não investigou explicitamente os efeitos de potenciais instabilidades físicas nas amostras de leite, como separação de fases ou formação de espuma, antes da aquisição espectral. Embora tenha sido realizada a homogeneização imediatamente antes da medição por inversão manual dos tubos Falcon, esse método pode não ter sido suficiente para eliminar todas as fontes de heterogeneidade. Outras não homogeneidades também podem ter surgido devido ao uso de placas de Petri sem comprimento de caminho óptico fixo, apesar da adição consistente de 10 g de cada amostra. No entanto, placas de Petri têm sido amplamente utilizadas com sucesso como suportes de amostra na espectroscopia NIR para diversas matrizes alimentares (Ji, Zhu, Deng, Jiang, & Chen, 2024; Ramo et al., 2024; Simeone et al., 2024; Oliveira, Badaró, Esquerre, Kamruzzaman, & Barbin,

2023; McVey, Gordon, Haughey, & Elliott, 2021; Srinuttrakul et al., 2021; Aykas et al, 2020; Kucha, & Ngadi, 2020; Badaró et al., 2019), incluindo leite de vaca e de cabra (Pereira et al., 2021; Pereira et al., 2020; de Lima, Andrade, da Silva, & Honorato, 2018). Essas aplicações abrangem tanto avaliações qualitativas, como técnicas de reconhecimento de padrões, quanto análises quantitativas usando modelos de calibração multivariada. Além disso, esses estudos demonstraram que a composição do material da placa de Petri (seja vidro ou poliestireno) não impacta significativamente os dados espectrais quando medições de branco (referência) apropriadas são realizadas.

Análises comparativas envolvendo espectrômetros de bancada e portáteis, utilizando placas de Petri idênticas, também mostraram que variações na resolução instrumental não afetam de maneira significativa a acurácia ou confiabilidade dos dados analíticos obtidos (Ramo et al., 2024; Oliveira, Badaró, Esquerre, Kamruzzaman, & Barbin, 2023; McVey, Gordon, Haughey, & Elliott, 2021; Srinuttrakul et al., 2021; Badaró et al., 2019; de Lima, Andrade, da Silva, & Honorato, 2018). Esses resultados são consistentes com o desempenho do método proposto no presente estudo, conforme evidenciado pelas figuras de mérito discutidas anteriormente. Apesar dessas considerações, estudos futuros devem adotar protocolos mais rigorosos para controlar a homogeneidade das amostras e investigar os efeitos das condições de armazenamento na integridade dos espectros. Esses esforços são particularmente importantes para o avanço da aplicação de técnicas espectroscópicas não destrutivas na detecção de adulteração de leite de cabra em condições reais de manuseio e armazenamento.

6 CONCLUSÕES

A metodologia proposta combina as vantagens de um espectrômetro de infravermelho próximo (NIR) miniaturizado com o Algoritmo de Projeções Sucessivas (SPA) para seleção de variáveis, resultando em melhorias significativas na precisão de predição para a análise de qualidade do leite de vaca. Comparado com um espectrômetro NIR de bancada, o dispositivo miniaturizado oferece maior portabilidade, custo-benefício e o potencial para análise no local, o que é particularmente vantajoso para aplicações que exigem tomada de decisões rápidas, como a detecção de adulteração do leite de cabra. Quando combinado com abordagens de calibração multivariada baseadas em SPA, a metodologia reduz os erros relativos de predição em 3,7% para quantificação de gordura e 3,2% para detecção de adulteração com leite de vaca. Essas melhorias ressaltam a eficácia do SPA em lidar com desafios associados à redundância de variáveis e multicolinearidade, que frequentemente comprometem o desempenho de modelos multivariados. Além disso, a natureza compacta do espectrômetro NIR miniaturizado aumenta sua aplicabilidade em ambientes de campo, superando as limitações logísticas e operacionais impostas pelos espectrofotômetros de bancada tradicionais. Essa sinergia metodológica representa um avanço significativo no desenvolvimento de ferramentas eficientes, precisas e práticas para avaliar a qualidade do leite de cabra, alinhando-se aos princípios de Química Verde na análise de alimentos para promover práticas mais sustentáveis e ambientalmente amigáveis.

Para as perspectivas futuras de continuidade deste trabalho, pretende-se ainda investigar mecanismos que possibilitem a quantificação do teor de adulteração em leite de cabra em pó, resultante da adição de adulterantes como extrato de soja, composto lácteo e leite de vaca, utilizando para isso imagens digitais associadas à calibração multivariada.

REFERÊNCIAS

- AYKAS, D. P.; BALL, C.; SIA, A.; ZHU, K.; SHOTTS, M. L.; SCHMENK, A.; RODRIGUEZ-SAONA, L. **In-situ screening of soybean quality with a novel handheld near-infrared sensor.** *Sensors*, [S. l.], v. 20, n. 21, p. 6283, 2020. DOI: <https://doi.org/10.3390/s20216283>.
- BADARÓ, A. T.; MORIMITSU, F. L.; FERREIRA, A. R.; CLERICI, M. T. P. S.; BARBIN, D. F. **Identification of fiber added to semolina by near infrared (NIR) spectral techniques.** *Food Chemistry*, [S. l.], v. 289, p. 195–203, 2019. DOI: <https://doi.org/10.1016/j.foodchem.2019.03.057>.
- BATTESTIN, C.; ÂNGELO, F. F.; ROCHA, R. A.; ANJOS, V.; BELL, M. J. V. **High resolution Raman spectroscopy of raw and UHT bovine and goat milk.** *Measurement: Food*, v. 6, p. 100029, 2022. doi.org/10.1016/j.meafoo.2022.100029.
- BÊC, K. B.; GRABSKA, J.; HUCK, C. W. **Miniaturized NIR Spectroscopy in Food Analysis and Quality Control: Promises, Challenges, and Perspectives. Review.** *Institute of Analytical Chemistry and Radiochemistry, University of Innsbruck, Innsbruck, Áustria*, 2022. doi.org/10.1016/j.foodcont.2022.107256.
- BRASIL.** Ministério da Agricultura e do Abastecimento. Instrução Normativa nº 37, de 31 de outubro de 2000. Aprova o regulamento técnico de identidade e qualidade do leite de cabra. Diário Oficial da União: seção 1, Brasília, DF, 8 nov. 2000. p. 16.
- BRASIL.** Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa nº 51, de 18 de setembro de 2002. Aprova o regulamento técnico de produção, identidade e qualidade do leite tipo A, do leite cru refrigerado, do leite pasteurizado e da coleta de leite cru refrigerado. Diário Oficial da União: seção 1, Brasília, DF, 20 set. 2002. p. 13.
- CANOVA, A.; SILVA, J. P. D.; CASTRO, M. A. DE; TORRES, CARLOS E. B.; GOMES, L. R. **An improved successive projections algorithm version to variable selection in multiple linear regression.** *Computers & Chemical Engineering*, [S.l.], v. 174, p. 107040, 2023. DOI: 10.1016/j.compchemeng.2023.107040.
- CEN, H.; HE, Y. **Theory and application of near infrared reflectance spectroscopy in determination of food quality.** *Trends in Food Science & Technology*, [S.l.], v. 18, n. 2, p. 72–83, fev. 2007. DOI: 10.1016/j.tifs.2006.09.003.
- CHEN, H.; TAN, C.; LIN, Z.; WU, T. **Classification and quantitation 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. DOI: 10.1016/j.saa.2017.08.034.
- CHU, C. *et al.* **Possible alternatives: identifying and quantifying adulteration in buffalo, goat, and camel milk using mid-infrared spectroscopy combined with modern statistical machine learning methods.** *Foods*, v. 12, p. 3856, 2023. doi.org/10.3390/foods12203856.
- 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. doi.org/10.3168/jds.2017-13287.

- DE LIMA, G. F.; ANDRADE, S. A. C.; DA SILVA, V. H.; HONORATO, F. A. **Multivariate classification of UHT milk as to the presence of lactose using benchtop and portable NIR spectrometers.** *Food Analytical Methods*, [S. l.], v. 11, n. 10, p. 2699–2706, 2018. DOI: <https://doi.org/10.1007/s12161-018-1253-7>.
- 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. DOI: 10.1039/C5AY00063C.
- DU, C. *et al.* **Single- and multiple-adulterants determinations of goat milk powder by NIR spectroscopy combined with chemometric algorithms.** *Agriculture*, v. 12, n. 3, p. 434, 2022. DOI 10.3390/agriculture12030434
- DU, C. *et al.* **Identification and quantification of goat milk adulteration using mid-infrared spectroscopy and chemometrics.** *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, v. 324, p. 124969, 2025. DOI 10.1016/j.saa.2024.124969.
- ESBENSEN, K. H.; GELADI, P.; LARSEN, A. **The RPD myth.** *NIR News*, [S. l.], v. 25, p. 24–28, 2014. DOI: <https://doi.org/10.1255/nirn.1462>.
- ETEMADI, S.; KHASHEI, M. **Etemadi multiple linear regression.** *Measurement*, v. 186, p. 110080, 2021. DOI 10.1016/j.measurement.2021.110080.
- FERREIRA, J. L. A.; ALMEIDA, L. F. A.; SIMÕES, S. S.; DINIZ, P. H. G. D.; FERNANDES, D. D. S. **Raman spectroscopy-based authentication of powder goat milk adulteration with cow milk.** *Food Control*, v. 167, p. 110800, 2025. DOI: 10.1016/j.foodcont.2024.110800.
- FERREIRA, M. M. C. **Quimiometria: conceitos, métodos e aplicações.** Campinas, SP: Editora da Unicamp, 2015. Disponível em: <https://doi.org/10.7476/9788526814714>.
- FODOR, M.; MATKOVITS, A.; BENES, E. L.; JÓK, Z. **The Role of Near-Infrared Spectroscopy in Food Quality Assurance: A Review of the Past Two Decades.** *Foods*, v. 13, p. 3501, 2024. DOI: 10.3390/foods13113501.
- FOLI, L. P.; HESPANHOL, M. C.; CRUZ, K. A. M. L.; PASQUINI, C. **Miniaturized Near-Infrared spectrophotometers in forensic analytical science – a critical review.** *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, v. 315, p. 124297, 2024. DOI: 10.1016/j.saa.2024.124297.
- GIUSSANI, B.; GORLA, G.; RIU, J. **Analytical chemistry strategies in the use of miniaturized NIR instruments: An overview.** *Critical Reviews in Analytical Chemistry*, [S. l.], v. 54, n. 1, p. 11–43, 2022. DOI: <https://doi.org/10.1080/10408347.2022.2047607>.
- GOMES, A. A.; AZCARATE, S. M.; DINIZ, P. H. G. D.; FERNANDES, D. D. S.; VERAS, G. **Variable selection in the chemometric treatment of food data: A tutorial review.** *Food Chemistry*, v. 370, p. 131072, 2022. doi.org/10.1016/j.foodchem.2021.131072
- GOMES, A. A.; GALVÃO, R. K. H.; DE ARAÚJO, M. C. U.; VÉRAS, G.; SILVA, E. C.

The successive projections algorithm for interval selection in PLS. *Microchemical Journal*, v. 110, p. 202–208, 2013. doi.org/10.1016/j.microc.2013.03.015

GRABSKA, J.; BEĆ, K. B.; HUCK, C. W. **Portability of miniaturized food analytical systems 4.0.** In: **HASSOUN, A. (Ed.).** *Developments in food quality and safety, food industry 4.0.* Academic Press, 2024. p. 189–231. DOI: <https://doi.org/10.1016/B978-0-443-15516-1.00010-4>.

HE, Y.; ZENG, W.; ZHAO, Y.; ZHU, X.; WAN, H.; ZHANG, M.; LI, Z. **Rapid detection of adulteration of goat milk and goat infant formulas using near-infrared spectroscopy fingerprints.** *International Dairy Journal*, v. 137, p. 105536, 2023. DOI: 10.1016/j.idairyj.2022.105536.

JL, Z.; ZHU, J.; DENG, J.; JIANG, H.; CHEN, Q. **Quantitative determination of zearalenone in wheat by the CSA-NIR technique combined with chemometric algorithms.** *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, [S. l.], v. 323, p. 124858, 2024. DOI: <https://doi.org/10.1016/j.saa.2024.124858>.

KHEMET, B.; RICHMAN, R. **A univariate and multiple linear regression analysis on a national fan (de)Pressurization testing database to predict airtightness in houses.** *Building and Environment*, v. 139, p. 1–12, 2018. DOI: 10.1016/j.buildenv.2018.09.030.

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 Part A: Molecular and Biomolecular Spectroscopy*, 189, 300–306, 2018. doi.org/10.1016/j.saa.2017.08.046

KUCHA, C. T.; NGADI, M. O. **Rapid assessment of pork freshness using miniaturized NIR spectroscopy.** *Journal of Food Measurement and Characterization*, [S. l.], v. 14, p. 1105–1115, 2020. DOI:10.1007/s11694-019-00360-9.

LESKINEN, H.; VENTTO, L.; KAIRENIUS, P.; SHINGFIELD, K. J.; VILKKI, J. **Temporal changes in milk fatty acid composition during diet-induced milk fat depression in lactating cows.** *Journal of Dairy Science*, [S. l.], v. 102, p. 1–13, 2019. DOI: <https://doi.org/10.3168/jds.2018-15860>.

LIAO, G.; WANG, T.; LI, X.; GU, J.; JIA, Q.; WANG, Z.; LI, H.; QIAN, Y.; QIU, J. **Comparison of the lipid composition of milk fat globules in goat (*Capra hircus*) milk during different lactations and human milk.** *Foods, Basel*, v. 13, n. 11, p. 1–15, 2024. DOI:10.3390/foods13111618.

LI, W.; HUANG, W.; FAN, D.; GAO, X.; ZHANG, X.; MENG, Y.; LIU, T. C. **Rapid quantification of goat milk adulteration with cow milk using Raman spectroscopy and chemometrics.** *Analytical Methods*, 15, 685–461, 2023. doi.org/10.1039/D2AY01697D.

LIRA, L. F. B. de; ALBUQUERQUE, M. S. de; PACHECO, J. G. A.; FONSECA, T. M.; CAVALCANTI, E. H. de S.; STRAGEVITCH, L.; PIMENTEL, M. F. **Infrared spectroscopy and multivariate calibration to monitor stability quality parameters of**

biodiesel. *Microchemical Journal*, v. 96, n. 1, p. 126-131, 2010. DOI: 10.1016/j.microc.2010.02.014.

MARIANI, N. C. T.; TEIXEIRA, G. H. A.; LIMA, K. M. G.; MORGENSTERN, T. B.; NARDINI, V.; CUNHA JÚNIOR, L. C. **NIRS and iSPA-PLS for predicting total anthocyanin content in jaboticaba fruit.** *Food Chemistry*, Amsterdam, v. 174, p. 643–648, 2015. DOI: 10.1016/j.foodchem.2014.11.008

MCVEY, C.; GORDON, U.; HAUGHEY, S. A.; ELLIOTT, C. T. **Assessment of the analytical performance of three near-infrared spectroscopy instruments (benchtop, handheld and portable) through the investigation of coriander seed authenticity.** *Foods*, [S. l.], v. 10, n. 5, p. 956, 2021. DOI: <https://doi.org/10.3390/foods10050956>.

MEHMOOD, T.; HOVDE LILAND, K.; SNIPEN, L.; SÆBØ, S. **A review of variable selection methods in partial least squares regression.** *Chemometrics and Intelligent Laboratory Systems*, v. 118, p. 62-69, 2012. DOI: 10.1016/j.chemolab.2012.05.001.

MOLSKA, I.; ZMARLICKI, S.; KONARZEWSKA, G.; JANICKI, A. **Effect of heat treatment on microbiological and keeping quality of milk stored at 4°C±2°C.** *Acta Alimentaria Polonica*, [S. l.], v. 15, n. 1, p. 25–34, 1989. Disponível em: <https://bibliotekanauki.pl/articles/1399527.pdf>.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to Linear Regression Analysis.** 5. ed. Tempe: Arizona State University, 2012.

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*, [S. l.], v. 77, p. R118–R126, 2012. DOI: <https://doi.org/10.1111/j.1750-3841.2012.02657.x>.

MUSA, M. A.; YANG, S. **Detection and quantification of cow milk adulteration using portable near-infrared spectroscopy combined with chemometrics.** *African Journal of Agricultural Research*, v. 17, p. 198-207, 2021. DOI: 10.5897/AJAR2020.15321.

NASCIMENTO, D. S.; INSAUSTI, M.; BAND, B. S. F.; LEMOS, S. G. **Simultaneous determination of Cu, Pb, Cd, Ni, Co and Zn in bioethanol fuel by adsorptive stripping voltammetry and multivariate linear regression.** *Fuel*, Amsterdam, v. 137, p. 172–178, 2014. doi.org/10.1016/j.fuel.2014.07.100

NAYIK, G. A.; JAGDALE, Y. D.; GAIKWAD, S. A.; DEVKATTE, A. N.; DAR, A. H.; ANSARI, M. J. **Nutritional profile, processing and potential products: A comparative review of goat milk.** *Dairy*, v. 3, n. 3, p. 622-647, 2022. DOI: 10.3390/dairy3030044.

OLIVEIRA, M. M.; BADARÓ, A. T.; ESQUERRE, C. A.; KAMRUZZAMAN, M.; BARBIN, D. F. **Handheld and benchtop vis/NIR spectrometer combined with PLS regression for fast prediction of cocoa shell in cocoa powder.** *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, [S. l.], v. 298, p. 122807, 2023. DOI: <https://doi.org/10.1016/j.saa.2023.122807>.

OLIVIERI, A. C. **Introduction to Multivariate Calibration: A Practical Approach.** 1. ed.

Cham: Springer Nature Switzerland AG, 2018. ISBN 978-3-319-97096-7. ISBN 978-3-319-97097-4 (eBook). DOI: 10.1007/978-3-319-97097-4.

O'MAHONY, J. A.; FOX, P. F. Chapter 2 - **Milk: an overview**. In: SINGH, H.; BOLAND, M.; THOMPSON, A. (Eds.). *Food Science and Technology: Milk Proteins*. 2. ed. Academic Press, 2014. p. 19-73. DOI: 10.1016/B978-0-12-405171-3.00002-7.

ORDÓÑEZ, J. A. **Tecnologia de alimentos: alimentos de origem animal**, v. 1. Porto Alegre: Artmed, 2005. 33p.

ORDÓÑEZ, J. A. **Tecnologia de alimentos: alimentos de origem animal**, v. 2. Porto Alegre: Artmed, 2005. 279p.

PARK, S. K.; MOON, H. J.; MIN, K. C.; HWANG, C.; KIM, S. **Application of a multiple linear regression and an artificial neural network model for the heating performance analysis and hourly prediction of a large-scale ground source heat pump system**. *Energy and Buildings*, v. 165, p. 206-215, 15 abr. 2018. DOI: 10.1016/j.enbuild.2018.01.029.

PASQUINI, C. **Near Infrared Spectroscopy: A Mature Analytical Technique with New Perspectives – A Review**. *Analytica Chimica Acta*, v. 1026, p. 8-36, 2018. DOI: 10.1016/j.aca.2018.04.004.

PASQUINI, C. **Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications**. *Journal of the Brazilian Chemical Society*, v. 14, n. 2, p. 198-219, 2003. DOI: 10.1590/S0103-50532003000200006.

PATIL, G. B. *et al.* **Milk adulteration: From detection to health impact**. *Food and Humanity*, v. 3, p. 100339, 2024. DOI: 10.1016/j.foohum.2024.100339.

PEREIRA, E. V. S.; FERNANDES, D. D. S.; ALMEIDA, L. F. A.; MACIEL, M. I. S.; DINIZ, P. H. G. D. **Goat milk authentication by one-class classification of digital image-based fingerprint signatures: Detection of adulteration with cow milk**. *Microchemical Journal*, v. 180, p. 107640, 2022. doi.org/10.1016/j.microc.2022.107640.

PEREIRA, E. V. S.; FERNANDES, D. D. S.; ARAÚJO, M. C. U.; DINIZ, P. H. G. D.; MACIEL, M. I. S. **In-situ authentication of goat milk in terms of its adulteration with cow milk using a low-cost portable NIR spectrophotometer**. *Microchemical Journal*, v. 163, p. 105885, 2021. doi.org/10.1016/j.microc.2020.105885.

PEREIRA, E. V. S.; FERNANDES, D. D. S.; ARAÚJO, M. C. U.; DINIZ, P. H. G. D.; MACIEL, M. I. S. **Simultaneous determination of goat milk adulteration with cow milk and their fat and protein contents using NIR spectroscopy and PLS algorithms**. *LWT*, v. 127, p. 109427, 2020. doi.org/10.1016/j.lwt.2020.109427.

PEREIRA, E. V. S. **Novas estratégias analíticas baseadas em espectroscopia no infravermelho próximo e imagens digitais para identificação e quantificação de adulterações em leite caprino**. 2020. Tese (Doutorado em Ciência e Tecnologia de Alimentos) – Universidade Federal da Paraíba, Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos, João Pessoa, 2020.

PINHEIRO, J. G. **Características físico-químicas do leite caprino na época seca e chuvosa na microrregião de Mossoró-RN**. 2012. 112 f. Dissertação (Mestrado em Produção Animal) — Universidade Federal Rural do Semiárido, Mossoró, 2012.

POMERANTSEV, A. L.; RODIONOVA, O. Y. **Multiclass partial least squares discriminant analysis: Taking the right way—A critical tutorial**. *Journal of Chemometrics*, [S.l.], v. 94, p. 1–11, 2018. DOI:10.1002/cem.3076.

RAMO, L. B.; NOBREGA, R. O.; FERNANDES, D. D. S.; LYRA, W. S.; DINIZ, P. H. G. D.; ARAUJO, M. C. U. **Determination of moisture, total protein, and phosphorus contents in powdered chicken egg samples using digital images, NIR spectra, data fusion, and multivariate calibration**. *Journal of Food Composition and Analysis*, [S. l.], v. 127, p. 105940, 2024. DOI: <https://doi.org/10.1016/j.jfca.2023.105940>.

RATH, S.; TRIPATHY, A.; TRIPATHY, A. R. **Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model**. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, v. 14, n. 5, p. 1467-1474, 2020. DOI: doi.org/10.1016/j.dsx.2020.07.045.

RICO, D. E.; HARVATINE, K. J. **Induction of and recovery from milk fat depression occurs progressively in dairy cows switched between diets that differ in fiber and oil concentration**. *Journal of Dairy Science*, [S. l.], v. 96, p. 6621–6630, 2013. DOI: <https://doi.org/10.3168/jds.2013-6820>.

RINNAN, A.; VAN DEN BERG, F.; ENGELSEN, S. B. **Review of the most common pre-processing techniques for near-infrared spectra**. *TrAC Trends in Analytical Chemistry*, v. 28, n. 10, p. 1201–1222, 2009. DOI: [10.1016/j.trac.2009.07.007](https://doi.org/10.1016/j.trac.2009.07.007).

SEN, S.; DUNDAR, Z.; UNCU, O.; ÖZEN, B. **Potential of Fourier-transform infrared spectroscopy in adulteration detection and quality assessment in buffalo and goat milks**. *Microchemical Journal*, v. 166, 106207, 2021. DOI: doi.org/10.1016/j.microc.2021.106207.

SILANIKOVE, N.; LEITNER, G.; MERIN, U.; PROSSER, C. G. **Goat milk composition and nutritional value: A review**. *Small Ruminant Research*, v. 89, n. 2-3, p. 110-124, 2010. DOI: [10.1016/j.smallrumres.2009.12.033](https://doi.org/10.1016/j.smallrumres.2009.12.033).

SIMEONE, M. L. F.; PIMENTEL, M. A. G.; QUEIROZ, V. A. V.; SANTOS, F.; BRITO, A.; AQUINO, L. F. M.; DA COSTA FILHO, J. C. E.; DE MENEZES, C. B.; PAES, M. C. D.; TIBOLA, C. S.; GUIMARÃES, P. E. O.; TRINDADE, R. S. **Portable near-infrared (NIR) spectroscopy and multivariate calibration for reliable quality control of maize and sorghum grain chemical composition**. *Journal of Food Composition and Analysis*, [S. l.], v. 134, p. 106502, 2024. DOI: <https://doi.org/10.1016/j.jfca.2024.106502>.

SINGH, A. K.; KUMAR, M.; SINGH, R. **The nutritional therapeutic values and chemical properties of goat milk**. *Journal of Nutrition, Metabolism and Health Sciences*, v. 5, n. 3, p. 118-123, 2022. DOI: [10.18231/j.ijnmhs.2022.022](https://doi.org/10.18231/j.ijnmhs.2022.022)

SKOOG, D. A.; HOLLER, F. J.; CROUCH, S. R. **Principles of Instrumental Analysis**. 6^a

ed. Belmont: Brooks Cole, 2009.

SOARES, S. F. C. **Um novo critério para seleção de variáveis usando o Algoritmo das Projeções Sucessivas**. 2010. 108 f. Dissertação (Mestrado em Química) – Universidade Federal da Paraíba, João Pessoa, 2010.

SRINUTTRAKUL, W.; MIHAILOVA, A.; ISLAM, M. D.; LIEBISCH, B.; MAXWELL, F.; KELLY, S. D.; CANNAVAN, A. **Geographical differentiation of Hom Mali rice cultivated in different regions of Thailand using FTIR-ATR and NIR spectroscopy**. *Foods*, [S. l.], v. 10, n. 8, p. 1951, 2021. DOI: <https://doi.org/10.3390/foods10081951>.

SUN, X.; ZHANG, Y.; ZHANG, B.; ZHANG, W.; XU, C. **Detection of fat content in peanut kernels based on chemometrics and hyperspectral imaging technology**. *Infrared Physics & Technology*, [S.l.], v. 107, p. 103261, 2020. DOI: 10.1016/j.infrared.2020.103226.

TANG, Q.; HUANG, L.; PAN, Z. **Multiple linear regression model for vascular aging assessment based on radial artery pulse wave**. *European Journal of Integrative Medicine*, v. 28, p. 92-97, 2019. DOI: 10.1016/j.eujim.2019.05.006.

TEIXEIRA, J. L. P.; CARAMÊS, E. T. S.; BAPTISTA, D. P.; GIGANTE, M. L.; PALLONE, J. A. L. **Vibrational spectroscopy and chemometrics tools for authenticity and improvement the safety control in goat milk**. *Food Control*, v. 112, p. 107105, 2020. DOI: 10.1016/j.foodcont.2020.107105.

TEIXEIRA, J. L. P.; CARAMÊS, E. T. S.; BAPTISTA, D. P.; GIGANTE, M. L.; PALLONE, J. A. L. **Rapid adulteration detection of yogurt and cheese made from goat milk by vibrational spectroscopy and chemometric tools**. *Journal of Food Composition and Analysis*, v. 96, p. 103712, 2021. DOI: 10.1016/j.jfca.2020.103712.

TEIXEIRA, J. L. P.; CARAMÊS, E. T. S.; BAPTISTA, D. P.; GIGANTE, M. L.; PALLONE, J. A. L. **Adulteration detection in goat dairy beverage through NIR spectroscopy and DD-SIMCA**. *Food Analytical Methods*, v. 15, p. 783–791, 2022. DOI: 10.1007/s12161-021-02151-9.

VALLESE, F. D. *et al.* **Exploiting the successive projections algorithm to improve the quantification of chemical constituents and discrimination of botanical origin of Argentinean bee-pollen**. *Journal of Food Composition and Analysis*, v. 126, p. 105925, 2024. DOI: 10.1016/j.jfca.2023.105925.

VERRUCK, S.; DANTAS, A.; PRUDENCIO, E. S. **Functionality of the components from goat's milk, recent advances for functional dairy products development and its implications on human health**. *Journal of Functional Foods*, v. 52, p. 243–257, 2019. DOI: 10.1016/j.jff.2018.11.017.

WANG, Y.; XIAO, R.; LIU, S.; WANG, P.; ZHU, Y.; NIU, T.; CHEN, H. **The impact of thermal treatment intensity on proteins, fatty acids, macro/micro-nutrients, flavor, and heating markers of milk—A comprehensive review**. *International Journal of Molecular Sciences*, [S. l.], v. 25, n. 16, p. 8670, 2024. DOI: <https://doi.org/10.3390/ijms25168670>.
WEI, W.; PENG, Y.; LI, Y.; QIAO, L. **Lightweight portable nondestructive detection technique for assessing meat freshness attributes based on light emitting diode array**. In:

PROCEEDINGS OF THE ASABE ANNUAL INTERNATIONAL MEETING, New Orleans, LA, USA, 26-29 jul. 2015. St. Joseph, MI: American Society of Agricultural and Biological Engineers, 2015. p. 1-6. doi:10.13031/aim.20152189624

WEI, W.; PENG, Y.; QIAO, L. **Development of hand-held nondestructive detection device for assessing meat freshness**. In: SENSING FOR AGRICULTURE AND FOOD QUALITY AND SAFETY VIII. Proceedings of the SPIE Commercial + Scientific Sensing and Imaging, Baltimore, MD, USA, 17-21 abr. 2016. Bellingham, WA: SPIE, 2016. p. 98640W. doi.org/10.1117/12.2223300

WIKING, L.; GREGERSEN, S. B.; HANSEN, S. F.; HAMMERSHØJ, M. **Heat-induced changes in milk fat and milk fat globules and its derived effects on acid dairy gelation: A review**. *International Dairy Journal*, [S. l.], v. 127, p. 105213, 2022. DOI: <https://doi.org/10.1016/j.idairyj.2021.105213>.

YADAV, A. K.; SINGH, J.; YADAV, S. K. **Composition, nutritional and therapeutic values of goat milk: a review**. *Asian Journal of Dairy and Food Research*, v. 35, n. 2, p. 96-102, 2016. DOI: 10.18805/ajdfr.v35i2.10719.

YAMAN, H. **A rapid method for detection adulteration in goat milk by using vibrational spectroscopy in combination with chemometric methods**. *Journal of Food Science and Technology*, v. 57, p. 3091–3098, 2020. DOI: 10.1007/s13197-020-04342-4.

YAZGAN, N. N.; GENIS, H. E.; BULAT, T.; TOPCU, A.; DURNA, S.; YETISEMIYEN, A.; BOYACI, I. H. **Discrimination of milk species using Raman spectroscopy coupled with partial least squares discriminant analysis in raw and pasteurized milk**. *Journal of the Science of Food and Agriculture*, v. 100, p. 4756–4765, 2020. DOI: 10.1002/jsfa.10534.

ZHAO, X.; WANG, Y.; LIU, X.; JIANG, H.; ZHAO, Z.; NIU, X.; LI, C.; PANG, B.; LI, Y. **Single- and multiple-adulterants determinations of goat milk powder by NIR spectroscopy combined with chemometric algorithms**. *Agriculture*, v. 12, n. 3, p. 434, 2022. DOI: 10.3390/agriculture12030434.