



UNIVERSIDADE
ESTADUAL DE LONDRINA

FABRÍCIO BRUNELI SILVA

**ESPECTROSCOPIA DO INFRAVERMELHO PRÓXIMO PARA
A DISCRIMINAÇÃO DAS ESPÉCIES DE CAFÉ CRU**

FABRÍCIO BRUNELI SILVA

**ESPECTROSCOPIA DO INFRAVERMELHO PRÓXIMO PARA
A DISCRIMINAÇÃO DAS ESPÉCIES DE CAFÉ CRU**

Dissertação apresentada à Universidade Estadual de Londrina - UEL, como requisito parcial para a obtenção do título de Mestre, em Química.

Orientador: Profa. Dra. Suzana Lucy Nixdorf.

Londrina
2026

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

Silva, Fabrício Bruneli.

ESPECTROSCOPIA DO INFRAVERMELHO PRÓXIMO PARA A DISCRIMINAÇÃO DAS ESPÉCIES DE CAFÉ CRU / Fabrício Bruneli Silva. - Londrina, 2026.
59 f.

Orientador: Suzana Lucy Nixdorf.
Dissertação (Mestrado em Química) - Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em Química, 2026.
Inclui bibliografia.

1. Quimiometria - Tese. 2. Fraude em café - Tese. 3. Café cru - Tese. 4. NIRS - Tese. I. Lucy Nixdorf, Suzana. II. Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa de Pós-Graduação em Química. III. Título.

CDU 54

FABRÍCIO BRUNELI SILVA

ESPECTROSCOPIA DO INFRAVERMELHO PRÓXIMO PARA A DISCRIMINAÇÃO DAS ESPÉCIES DE CAFÉ CRU

Dissertação de mestrado apresentada à Universidade Estadual de Londrina – UEL, como requisito parcial para a obtenção do título de Mestre em Química.

BANCA EXAMINADORA

Orientador: Profa. Dra. Suzana Lucy Nixdorf
Universidade Estadual de Londrina – UEL

Prof. Dr. Diego Galvan
Universidade Federal de Santa Catarina –
UFSC

Prof. Dr. Mario Henrique Montazzolli Killner
Universidade Estadual de Londrina – UEL

Londrina, 13 de março de 2026.

AGRADECIMENTOS

A Deus, por todas as maravilhas que tem feito na minha vida, concedendo-me força e persistência mesmo diante das dificuldades.

À minha orientadora e professora, Dra. Suzana Lucy Nixdorf, por todo o conhecimento compartilhado, pelo auxílio e pelos ensinamentos durante o desenvolvimento deste trabalho e ao longo de todo o mestrado.

À Banca Examinadora, pelo aceite e pelas valiosas contribuições a este trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo fomento e concessão da bolsa, fundamentais para o desenvolvimento desta pesquisa.

Ao Ministério da Agricultura e Pecuária (MAPA) e ao Instituto de Desenvolvimento Rural do Paraná – IAPAR, pela disponibilização das amostras de café analisadas neste trabalho. À Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA Instrumentação, em especial ao Dr. Ednaldo José Ferreira, pelos ensinamentos e pelo auxílio essencial durante o desenvolvimento deste estudo.

À Universidade Estadual de Londrina (UEL) e ao Laboratório de Apoio à Pesquisa Agropecuária (LAPA), que integra a Central Multiusuários de Pesquisa da UEL (CMLP), pelo suporte técnico e infraestrutura disponibilizados.

Aos amigos Thiago, Rafael, Beatriz, Paula e Francisnara, pelo apoio, parceria e por tornarem os dias mais leves e alegres.

A todos os integrantes do Grupo DIA (Desenvolvimento de Instrumentação, Automação e Metodologia Analítica), que contribuíram para o desenvolvimento deste trabalho, e aos professores coordenadores do Laboratório de Análise Físico-Química da Água (LAFQ), pelo acolhimento e apoio.

À minha mãe, Claudia, que sempre me incentivou e, com muito amor, esteve ao meu lado, apoiando-me em todas as etapas e fazendo o possível para que eu realizasse meus sonhos.

SILVA, Fabrício Bruneli. **Espectroscopia do infravermelho próximo para a discriminação das espécies de café cru**. 2026. 58 f. Dissertação (Programa de Mestrado em Química) – Universidade Estadual de Londrina, Londrina, 2026.

RESUMO

O café possui grande importância na economia do Brasil, sendo responsável por cerca de 30% da produção mundial. Em 2024, foram exportados aproximadamente 50,6 milhões de sacas de 60 kg. As espécies *Coffea arabica* e *Coffea canephora* são as mais comercializadas, sendo o arábica mais valorizado devido às suas características sensoriais, como sabor mais suave e adocicado, além das maiores exigências relacionadas ao seu cultivo. Em função dessas diferenças, o café arábica pode atingir valores de mercado significativamente superiores aos do canéfora, o que favorece a ocorrência de fraudes. Assim, o desenvolvimento de uma metodologia rápida, de baixo custo e robusta para a identificação da espécie apresenta grande relevância econômica. Nesse contexto, este trabalho teve como objetivo desenvolver uma metodologia baseada na espectroscopia no infravermelho próximo (NIRS) para discriminar as espécies arábica e canéfora, tanto em grão quanto moídas. No total, foram utilizadas 207 amostras provenientes das seis principais regiões produtoras de café do Brasil (MG, SP, BA, PR, ES e RO), referentes às safras de 2020 e 2022. As amostras da safra de 2022 foram utilizadas para a calibração dos modelos, enquanto as da safra de 2020 foram empregadas para a validação externa. Os espectros NIR foram obtidos a partir de 50 g de café em grão e 2,0 g de café moído, sendo este último passado em peneira de 32 mesh, na faixa de 400 a 2498 nm, por meio da média de 32 scans. Os espectros foram submetidos aos pré-processamentos de correção multiplicativa de espalhamento (MSC), variável padrão normal (SNV) e primeira derivada. A análise por componentes principais (PCA) indicou que as componentes principais 1 e 2 (PC1 e PC2), obtidas a partir dos espectros tratados com MSC, explicaram juntas 95,75% da variância para o café moído e 97,51% para o café em grão. A Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) foi empregada para a discriminação das duas espécies e para avaliar a eficiência da seleção de variáveis associadas a compostos presentes em diferentes concentrações nas duas espécies. Para isso, os modelos foram construídos utilizando o espectro NIR completo (1000 a 2498nm) e com regiões selecionadas. Para o café cru moído, o modelo apresentou exatidão de 99,34% na calibração e 100% na validação. Para o café em grão, observou-se alta eficiência na calibração (100%), porém baixo desempenho no conjunto de validação (18,87%). Os modelos com seleção de variáveis apresentaram desempenho semelhante ao obtido com o espectro completo. A partir das regiões selecionadas, foi utilizado um algoritmo iterativo com o objetivo de encontrar a melhor razão entre somas de comprimentos de onda capaz de discriminar as duas espécies, considerando como critérios a maximização da distância entre os grupos e a minimização da variabilidade dentro dos grupos. A equação obtida foi posteriormente ajustada visando maior explicabilidade química, por meio da inclusão de novos comprimentos de onda. Com a equação final, tornou-se possível discriminar as duas espécies de café provenientes de diferentes safras.

Palavras-chave: NIRS; quimiometria; discriminação de espécie; café cru; método rápido.

SILVA, Fabrício Bruneli. **Near-infrared spectroscopy for the discrimination of raw coffee species**. 2026. 58 p. Chemistry Master's Degree Dissertation – State University of Londrina, Londrina, 2026.

ABSTRACT

Coffee plays an important role in the Brazilian economy, accounting for approximately 30% of global production. In 2024, around 50.6 million 60-kg bags were exported. The species *Coffea arabica* and *Coffea canephora* are the most commercialized, with arabica being more highly valued due to its sensory characteristics, such as a smoother and sweeter flavor, as well as the greater agronomic requirements associated with its cultivation. As a result of these differences, arabica coffee can reach significantly higher market values than canephora, which favors the occurrence of fraud. Thus, the development of a fast, low-cost, and robust methodology for species identification is of great economic relevance. In this context, this study aimed to develop a methodology based on near-infrared spectroscopy (NIRS) to discriminate between arabica and canephora species, both in whole bean and ground forms. A total of 207 samples from the six main coffee-producing regions of Brazil (MG, SP, BA, PR, ES, and RO), corresponding to the 2020 and 2022 harvests, were analyzed. Samples from the 2022 harvest were used for model calibration, while those from 2020 were used for external validation. NIR spectra were obtained from 50 g of whole beans and 2.0 g of ground coffee, the latter sieved through a 32-mesh screen, in the spectral range from 400 to 2498 nm, using the average of 32 scans. The spectra were subjected to three preprocessing methods: multiplicative scatter correction (MSC), standard normal variate (SNV), and first derivative. Principal component analysis (PCA) showed that the first two principal components (PC1 and PC2), obtained from MSC-treated spectra, jointly explained 95.75% of the variance for ground coffee and 97.51% for whole beans. Partial Least Squares Discriminant Analysis (PLS-DA) was employed to discriminate between the two species and to evaluate the effectiveness of variable selection associated with compounds present at different concentrations in the two species. For this purpose, the models were built using the full NIR spectrum (1000–2498 nm) and selected spectral regions. For ground raw coffee, the model showed an accuracy of 99.34% in calibration and 100% in validation. For whole beans, high efficiency was observed in calibration (100%), but low performance was obtained in the validation set (18.87%). The models with variable selection showed similar performance to those obtained with the full spectrum. Based on the selected regions, an iterative algorithm was applied to find the best ratio between sums of wavelengths capable of discriminating the two species, considering as criteria the maximization of intergroup distance and the minimization of intragroup variability. The resulting equation was subsequently adjusted to improve chemical interpretability by including additional wavelengths. With the final equation, it became possible to discriminate between the two coffee species from different harvests.

Key-words: NIRS; chemometrics; species discrimination; raw coffee; rapid method.

LISTA DE FIGURAS

Figura 1 - Grãos secos de café arábica e café canéfora	15
Figura 2 - Imagem hiper espectral de café 100% Arábica (A), 100% Canéfora (B) e 50% arábica (esquerda) e 50% Canéfora (direita)	18
Figura 3 - Espectro NIRS das amostras de café cru em grão e moído	27
Figura 4 - Espectro NIRS das amostras de café cru moído e em grão corrigido com MSC.....	28
Figura 5 - Espectros NIRS das amostras de cafés arábica e canéfora crus em grãos e moído corrigido com SNV.....	29
Figura 6 - Espectros NIRS das amostras de cafés arábica e canéfora crus em grãos e moído corrigido com primeira derivada	30
Figura 7 - Gráfico de <i>scores</i> para as amostras de café cru-utilizando MSC, para o café moído (A) e em grão (B).....	31
Figura 8 - Gráfico de <i>scores</i> para as amostras de café cru-utilizando SNV, para o café moído (A) e em grão (B).....	33
Figura 9 - Gráfico de <i>scores</i> para as amostras de café cru-utilizando alisamento Savitzky Golay e primeira derivada, para o café moído (A) e em grão (B).....	34
Figura 10 - Gráfico de loadings da PC1, PC2 e PC3 para as amostras de café cru moído e em grão	36
Figura 11 - Respostas da PLS-DA para as amostras de café cru moído utilizando o espectro completo e com variáveis selecionadas	39
Figura 12 - Respostas da PLS-DA para as amostras de café cru em grão utilizando o espectro completo e com seleção de variáveis	40
Figura 13 - <i>Ratio</i> para as amostras de calibração e validação das amostras de café cru moído	44
Figura 14 - <i>Ratio</i> para as amostras de calibração e validação das amostras de café cru em grão	44
Figura 15 - <i>Ratio</i> para as amostras de calibração e validação das amostras de café cru moído	46
Figura 16 - <i>Ratio</i> para as amostras de calibração e validação das amostras de café cru em grão	46

Figura 17 - Gráfico de Scores da PCA utilizando o espectro visível para o café cru moído (A) e em grão(B).....	48
Figura 18 - Respostas da PLS-DA para as amostras de café cru moído e em grão utilizando o espectro visível	49
Figura A.1 - Resultado PLS-DA utilizando o espectro pré-processado com SNV para o café cru moído e em grão	56
Figura A.2 - Resultado PLS-DA utilizando o espectro pré-processado com primeira derivada para o café cru moído e em grão.....	57

LISTA DE TABELAS

Tabela 1 - Composição química das duas espécies de café cru	15
Tabela 2 - Bandas, atribuições e compostos relacionado aos comprimentos de onda das bandas selecionadas	38
Tabela 3 - Parâmetros de desempenho do PLS-DA para o café arábica cru moído	40
Tabela 4 - Matriz de confusão do PLS-DA para café moído classe arábica	40
Tabela 5 - Parâmetros de desempenho do PLS-DA para o café arábica cru em grão	41
Tabela 6 - Matriz de confusão do PLS-DA para café moído classe arábica	41
Tabela 7 - Comprimentos de onda do numerador e denominador para a separação das duas espécies em café moído	43
Tabela 8 - Comprimentos de onda do numerador e denominador para a separação das duas espécies em café cru em grão	43
Tabela 9 - Comprimento de onda do numerador e denominador da razão para o café cru moído da Equação modificada	45
Tabela 10 - Comprimento de onda do numerador e denominador da razão para o café cru em grão da Equação modificada	45
Tabela 11 - Parâmetros de desempenho do modelo baseado em uma Equação obtida por algoritmo e Equação modificada para café cru moído	47
Tabela 12 - Parâmetros de desempenho do modelo baseado em uma Equação obtida por algoritmo e Equação modificada para café cru em grão	47
Tabela A.1 - Matriz de confusão do PLS-DA utilizando espectro pré-processado com SNV para o café cru moído e em grão para classe arábica	58
Tabela A.2 - Matriz de confusão do PLS-DA utilizando espectro pré-processado com primeira derivada para o café cru moído e em grão para classe arábica	58

SUMÁRIO

1	INTRODUÇÃO	11
2	OBJETIVOS	13
2.1	Objetivos geral	13
2.2	Objetivos específicos	13
3	REVISÃO BIBLIOGRÁFICA	14
3.1	História do café	14
3.2	Importância econômica do café	16
3.3	Identificação de adulteração em café	17
3.4	Espectroscopia do infravermelho próximo	19
3.5	Quimiometria	21
4	MATERIAIS E MÉTODOS	24
4.1	Amostras	24
4.2	Obtenção dos espectros	24
4.3	Análise quimiométrica e softwares	25
5	RESULTADOS E DISCUSSÕES	27
5.1	Espectros infravermelho próximo	27
5.2	Análise por Componentes Principais (PCA)	31
5.3	Análise discriminante por mínimos quadrados parciais	37
5.4	Discriminação da espécie de café moído e em grão	42
5.5	Aplicação do espectro visível na discriminação das espécies de café	48
6	CONCLUSÃO	51
	REFERÊNCIAS	52
	APÊNDICE A	56

1 INTRODUÇÃO

O café é uma commodity agrícola importante mundialmente, sendo uma das mais comercializadas. Consumida mundialmente devido aos seus benefícios para a saúde humana ou devido as suas características sensoriais. Existem diversas espécies do gênero *Coffea*, onde as duas espécies mais produzidas e consumidas são a *Coffea arabica* e *Coffea canephora*. O cultivo de café sempre esteve presente o desenvolvimento econômico e histórico brasileiro. O Brasil sempre teve um importante papel na produção de café mundial, onde nas primeiras décadas de 1900, foi responsável por cerca de 75% da produção mundial, atuando diretamente na formação dos preços externos dessa importante commodity (Carvalho et al., 2017; Silva et al., 2023)

Atualmente, o café é uma das bebidas mais consumidas mundialmente, e continua sendo um produto importante para a economia brasileira, sendo o maior produtor mundial de café, o Brasil atende o mercado externo e o mercado interno, e é responsável por cerca de 30% de toda produção mundial (Wang, 2025).

De acordo com a Associação Brasileira da Indústria de Café (ABIC), em 2024, o Brasil produziu um total de 54,3 milhões de sacas de 60 kg, das quais, 50,6 milhões foram destinadas à exportação. Do total exportado, 91,8% correspondeu ao café cru, enquanto 8,2% referiam-se ao café torrado solúvel. A receita proveniente dessa exportação somou 12,6 bilhões de dólares, sendo que 11,8 bilhões foram gerados especificamente pelo café cru (ABIC, 2025a).

De todo o café exportado, cerca de 80% do volume é do *Coffea arabica*, enquanto o restante é do *Coffea canephora*. As diferenças entre as espécies vão desde a composição química, onde o arabica possui uma menor concentração de cafeína, e maior quantidade de sacarose, que impacta diretamente nas notas sensoriais, tornando-o uma bebida mais suave e doce, sendo essa preferível pelo mercado externo. Esse fator, associado a maior dificuldade de cultivo, faz com que esse café possua um maior valor quando comparado com o canephora (Cestari, 2021).

Essa diferença no preço, possibilita a ocorrência de fraude, na qual grãos de café canephora, são vendidos como arabica. A Instrução Normativa nº 8, de 11 de junho de 2003, do Ministério da Agricultura, Pecuária e Abastecimento (MAPA), define a “prova de xícara” como metodologia oficial na identificação dessa fraude. A Prova

de xícara é uma análise sensorial, realizada por provadores treinados, que por meio dos aromas e sabores, possibilitam a identificação da espécie do café cru analisado (BRASIL, 2003).

A análise sensorial é muito importante para a análise de produtos, com ela é possível identificar qualitativamente e quantitativamente os atributos sensoriais. Entretanto, ela possui algumas desvantagens, como a necessidade de provadores treinados capazes de identificar as nuances de sabores em produtos semelhantes. Sendo uma análise que é baseada na percepção humana, é uma análise subjetiva, assim provadores de diferentes regiões podem chegar a resultados diferentes para uma mesma amostra (Marques et al., 2022; Varela; Ares, 2012).

2 OBJETIVOS

2.1 OBJETIVOS GERAL

Desenvolver uma metodologia rápida e não destrutiva utilizando a espectroscopia do infravermelho próximo e quimiometria, capaz de conseguir discriminar a espécie de café cru brasileiro, moído e em grão, de diferentes safras.

2.2 OBJETIVOS ESPECÍFICOS

- Utilizar a análise por componentes principais para avaliar a capacidade de diferenciação entre as duas espécies de café cru e identificar as regiões do espectro correlacionadas a essa separação.
- Selecionar as regiões do espectro infravermelho correlacionando diferenças de composição química das duas espécies de café cru e avaliar por meio da análise de discriminante por mínimos quadrados parciais (PLS-DA) a efetividade da seleção das variáveis frente ao espectro completo.
- Utilizar os comprimentos de onda selecionados para encontrar uma razão entre comprimentos de onda que consiga discriminar as duas espécies de café cru e auxiliar o controle de qualidade.

3 REVISÃO BIBLIOGRÁFICA

3.1 HISTÓRIA DO CAFÉ

O café é uma planta etíope e atualmente ainda faz parte da vegetação natural. Os europeus foram os responsáveis pela disseminação do café pelo mundo. Existem diversas lendas em torno da descoberta do café, sendo a mais conhecida a Lenda de Kaldi, que foi registrada em manuscritos do lêmên do ano de 575. Kaldi era um pastor que observou que ao mastigar os frutos amarelo avermelhados dos arbustos abundantes dos campos da região suas cabras ficavam alegres e com mais energia. Todavia, foram os árabes que inicialmente dominaram o plantio e o preparo do café, onde era feito uma infusão com cerejas e os grãos de café fervidas em água. Essa planta foi denominada *Kaweh* e a bebida obtida a partir dela recebeu o nome de Kahwah ou Cahue (ABIC, 2025b).

O café foi introduzido no Brasil no ano de 1727 quando o oficial português Francisco de Mello Palheta trouxe as primeiras mudas de café e estas foram plantadas no Pará, florescendo com facilidade na região. No final do século 19 e início do século 20, o café se tornou o principal produto exportado, e foi responsável pelo desenvolvimento de grandes centros urbanos como São Paulo e Rio de Janeiro (ABIC, 2021).

Pertencente ao gênero *Coffea. L.* (Rubiácea), atualmente são conhecidas mais de 100 espécies de café, sendo os mais importantes o *Coffea arabica* (café arábica) e o *Coffea canephora* (Café canéfora), conhecida também como robusta, correspondem respectivamente a 70% e 30% da produção nacional (Zhang et al., 2020). Esses produzem frutos conhecidos como cerejas, no qual no interior de cada fruto contém dois invólucros, polpa e contém dois grãos de café. Os pés de café em geral podem atingir de 2 a 2,5 metros de altura, podendo chegar até 10 metros (Waters; Arendt; Moroni, 2017)

As duas espécies de cafés citadas anteriormente podem ser diferenciadas por meio de diversos fatores, sendo o café arábica cultivado em altitudes maiores com temperatura na faixa de 18 a 20 °C, em climas mais amenos. Este se reproduz por meio da autofecundação e apresenta um grão cru em formato mais ovalar, comumente comercializado em sua forma pura, ou em forma de *blend*. Em contrapartida, o café canéfora é produzido em alturas mais baixas, em temperaturas mais elevadas, sua

fecundação ocorre de forma cruzada, por meio de insetos polinizadores e ação dos ventos. Esse possui uma resistência maior a pragas e seu grão cru apresenta uma forma mais arredondada e coloração castanha (Kalschne et al., 2019; Mihailova et al., 2022)

Figura 1 - Grãos secos de café arábica e café canéfora



Coffea arabica



Coffea canephora

A composição química das duas espécies também é diferente, essas diferenças impactam diretamente nas características sensoriais das bebidas, cuja composição química do café cru está apresentada na Tabela 1.

Tabela 1 - Composição química das duas espécies de café cru

Componente	<i>Coffea arabica</i> ^a	<i>Coffea canephora</i> ^a
Polissacarídeos	43,0 - 45,0	46,9 - 48,3
Sacarose	6,2 - 8,5	0,9 - 4,8
Lipídeos	15 - 18	8 - 12
Proteínas	8,5 - 12	8,5 - 12
Ácidos clorogênicos	6,7 - 9,2	7,1 - 12,1
Cafeína	0,8 - 1,4	1,7 - 2,4
Trigonelina	0,6 - 1,2	0,3 - 0,9
Água	8 - 12	8 - 12

^aValores referentes à concentração de g 100 g⁻¹ em base seca

Fonte: (Wang; Lim, 2015)

A composição química do café é complexa, desse modo, diversos compostos químicos são encontrados no café que determinam características sensoriais e efeitos no corpo. Essas diferenças estão relacionadas a variedade do grão, a região onde o grão é cultivado, fatores ambientais, assim como o processo de torrefação e o tempo de maturação dos grãos(Wei; Tanokura, 2015)

Os principais constituintes do café são a cafeína, a trigonelina, os ácidos clorogênicos, açúcares, aminoácidos e minerais. Quando o café é torrado, a

trigonelina e os ácidos clorogênicos são responsáveis pelos aromas e sabores, enquanto a cafeína não é afetada por esse processo, sendo diretamente ligado aos efeitos farmacológicos do café (Atlabachew et al., 2021)

A cafeína é uma metil xantina, que são substâncias capazes de estimular o sistema nervoso, sendo considerada uma droga estimulante, que é comumente utilizada em bebidas energéticas e fármacos. Os efeitos notáveis ocorrem após a ingestão de pequenas a moderadas doses desse composto, onde é notável uma melhoria no desempenho psicomotor e cognitivo do consumidor. Entretanto, o consumo em excesso é prejudicial à saúde (Bervelieri Madeira et al., 2017)

As duas espécies de café não apresentam apenas diferenças botânicas e químicas, porém apresentam também diferenças no preço de mercado, onde o café arábica apresenta um valor superior ao café canéfora. Essa valorização ocorre devido ao arábica produzir uma bebida mais suave, sendo uma bebida com uma aceitação maior no paladar mundial, enquanto o café canéfora, apresenta uma bebida mais forte, com um amargor maior e maior adstringência, comumente é utilizado em *blends* (CESTARI, 2021; KAMILOGLU, 2019).

3.2 IMPORTÂNCIA ECONÔMICA DO CAFÉ

O café é um cultivar extremamente importante para a economia brasileira e sua exportação cresce anualmente, a qual no ano de 2024 o Brasil foi responsável por 50,6 milhões de sacas de 60 kg de café cru, sendo 46,4 milhões correspondem ao café cru, correspondendo a 31% do volume total de café cru exportado do mundo (ABIC, 2025b).

Visto que cerca de 70% da produção do café brasileiro é da espécie arábica, grande parte desse café é exportado, enquanto uma pequena parte é mantida no comércio interno, onde é comercializado a maior parte do café canéfora. No período de 2023 a 2024, a Associação Brasileira da Indústria de Café (ABIC) registrou um consumo de 21,9 milhões de sacas de café, representando um aumento de 1,1% com relação ao período anterior, que havia sido de 21,7 milhões de sacas. O consumo per capita foi de 6,26 kg de café cru e de 5,01 kg de café torrado e moído, superando o consumo de 2022, que foi de 5,96 kg e 4,77 kg respectivamente.

As principais áreas cafeeiras no Brasil estão localizadas nas regiões Centro-oeste e Sul, onde se destacam quatro dos maiores produtores: Minas Gerais, São Paulo, Espírito Santo e Paraná. As regiões Norte e nordeste apresentam lavouras de

café concentradas mais nos estados da Bahia e de Rondônia. Dentre os estados produtores, São Paulo, Minas Gerais, Paraná, Bahia e parte do Espírito Santo possuem lavouras que são dedicadas exclusivamente para a produção de café arábica, enquanto Rondônia e parte do Espírito Santo possuem áreas predominantemente dedicada a produção de café canéfora (Fornazier et al., 2017; Volsi et al., 2019).

Atualmente um dos principais problemas da indústria cafeeira é a adulteração do café, podendo ocorrer de três formas diferentes, podendo ela ser devido a adição de diversos grãos ao produto, como soja, arroz, milho, cevada, assim como cascas e paus de cafeeiro; pode ocorrer também por meio da adição de cafés de espécies mais baratas como o café canéfora em cafés arábicas, sendo vendido como arábica puro; e por meio da mistura de espécies de cafés de baixa qualidade com cafés de alta qualidade de diferentes regiões (Briandet; Kemsley; Wilson, 1996)

No Brasil, a portaria da secretaria de Defesa Agropecuária N° 570, de 9 de maio de 2022 estabelece parâmetros e limites para a presença de matérias estranhas e impurezas para que o café possa ser consumido. Desse modo, não é permitido a presença de nenhuma outra substância que não seja proveniente do cafeeiro, sendo permitido a presença de apenas 1% de cascas de café ou paus de cafeeiro.

A identificação da adulteração do café pode ser identificada por meio da análise sensorial de profissionais que passam por um treinamento, se tornando capaz de conseguir identificar a presença de misturas de diferentes espécies de café, todavia, a análise sensorial é uma forma subjetiva, visto que leva em conta a experiência anterior do provador, assim como suas preferências. Quando torrado e moído, a identificação da presença dessa adulteração se torna ainda mais difícil, podendo ser utilizado a microscopia para a sua identificação. Todavia, esta técnica é subjetiva e muitas vezes ineficiente nessa identificação, sendo necessário a utilização de outras técnicas com maior reprodutividade, precisão e sensibilidade para a identificação, como a cromatografia líquida de alta eficiência, a cromatografia gasosa, espectroscopia do infravermelho próximo, entre outras (Toci et al., 2016)

3.3 IDENTIFICAÇÃO DE ADULTERAÇÃO EM CAFÉ

As técnicas utilizadas para a identificação da adulteração podem ser tanto métodos químicos, quanto métodos físicos. Os métodos físicos, utilizam-se de imagens hiperespectral e multiespectral para a identificar adulteração de impurezas e

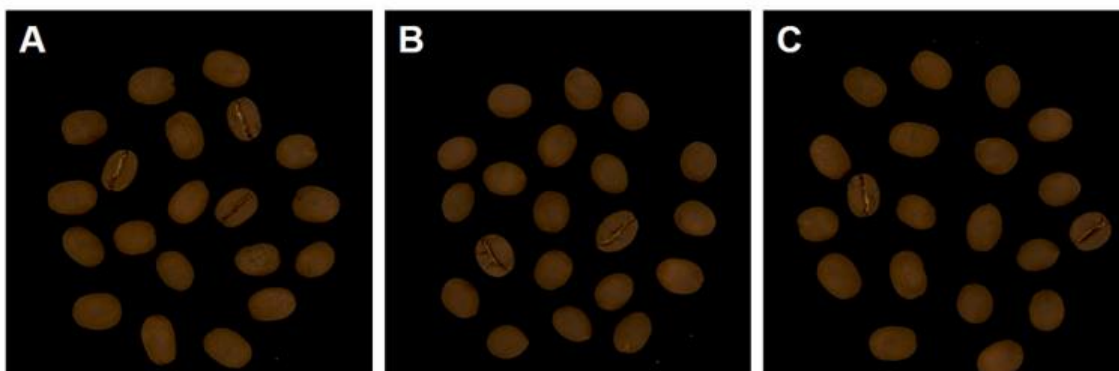
matérias estranhas, como a adição de espécies diferentes ao café arábica. Ambas as técnicas geram imagens de amostras (Figura 1) e com base em um treinamento prévio, com amostras feitas com adulteração conhecida, é possível fazer a diferenciação.

A técnica de imagens hiperespectrais permite a captura de dados em centenas de bandas espectrais contínuas, em vez de apenas três bandas tradicionais (vermelho, verde e azul). Esse método gera um "cubo" de dados onde cada camada representa um comprimento de onda específico, fornecendo um perfil espectral extenso para cada pixel da imagem.

Cada substância interage com a luz de maneira única, absorvendo, refletindo e emitindo luz em comprimentos de onda distintos. Essa interação resulta em um padrão espectral específico, que pode ser usado para identificar e distinguir entre diferentes materiais. No caso do café, os perfis espectrais dos grãos puros e dos adulterantes exibem características distintas, possibilitando a detecção de adulterações pela análise dos dados espectrais obtidos (Mihailova et al., 2022).

O processo de análise hiperespectral envolve o pré-processamento dos dados para a remoção de ruídos e correções atmosféricas, garantindo a precisão das informações. Em seguida, algoritmos avançados extraem características espectrais dos *pixels* e aplicam técnicas de aprendizado de máquina e métodos estatísticos, para comparar os perfis espectrais das amostras suspeitas com os perfis de referência de café puro e adulterado. A detecção de discrepâncias entre esses perfis ajuda a identificar a presença de adulterantes (Forchetti; Poppi, 2020; Mihailova et al., 2022)

Figura 2 - Imagem hiperespectral de café 100% Arábica (A), 100% Canéfora (B) e 50% arábica (esquerda) e 50% Canéfora (direita)



Fonte: Mihailova et al., 2022

Outra técnica física, que também é utilizada na identificação de adulteração de café é a espectroscopia do infravermelho, em que os espectros são obtidos e possibilitam verificar a presença de compostos que estão presentes no adulterante e não no café, atuando como 'marcadores'. Desse modo, a utilização de técnicas quimiométricas, como a análise por componentes principais (PCA) podem ser utilizadas para identificar a presença de adulterantes, e a regressão por mínimos quadrados parciais (PLS) podem ser utilizadas para quantificar essa adulteração. Além disso a criação de instrumentos NIR portáteis e com modelos PLS criados, é possível que a identificação e quantificação se torne portátil, auxiliando a fiscalização (Correia et al., 2018; Couto et al., 2022)

Para métodos químicos, uma técnica comumente utilizada é a cromatografia líquida de troca iônica com detector de pulso amperométrico. A cromatografia de troca aniônica (AEX) baseia-se na separação de íons dependendo de suas interações com uma resina carregada negativamente na coluna cromatográfica. Os íons são eluídos da amostra em diferentes tempos de retenção em relação à sua afinidade pela resina. Esta técnica de perfilagem de açúcares é bem estabelecida porque separa eficientemente açúcares simples de açúcares complexos (Sanders et al., 2017)

O detector por pulso amperométrico (PAD) oferece detecção altamente sensível e seletiva para compostos que são oxidados ou reduzidos eletroquimicamente. O PAD funciona aplicando um potencial cíclico ao eletrodo de trabalho, que alterna entre as fases de limpeza, oxidação e redução. Isso mantém a superfície do eletrodo limpa e melhora sua precisão (Natale et al., 2015)

Os carboidratos, em particular, são oxidados sob condições específicas, permitindo sua quantificação, sendo possível obter os perfis cromatográficos dos adulterantes, assim como dos cafés puros, possibilitando a identificação da adulteração (Pauli et al., 2014)

3.4 ESPECTROSCOPIA DO INFRAVERMELHO PRÓXIMO

A espectroscopia do infravermelho próximo (NIR) se baseia na aplicação de um feixe luminoso que é incidido sobre uma amostra, e com base na interação da luz com as ligações químicas dos compostos presentes na amostra, ocorre a vibração dessas ligações, sendo essas frequências relacionadas com o tipo da ligação (Cozzolino, 2021)

O espectro de infravermelho próximo pode ser obtido de duas maneiras diferentes, por transmitância, na qual o feixe de luz é direcionado sobre a amostra e o detector recebe o feixe que não foi absorvido pela amostra. Outro método que pode ser utilizado é por meio da reflectância, na qual o feixe luminoso é aplicado na amostra, e o feixe refletido da amostra é direcionada para o detector.

A espectroscopia do infravermelho próximo é uma fração do espectro do infravermelho e compreende a faixa de 750 a 2400 nm. Essa energia é mais que o necessário para promover moléculas apenas para o mais baixo nível de energia excitado, porém não é suficiente para promover a excitação eletrônica das moléculas (Pasquini, 2003)

Devido a menor energia, as informações obtidas em um espectro NIRS, está relacionado aos sobretons e combinações vibracionais das moléculas, sendo essas transições mais fracas e menos específicas. Essas características fazem com que a análise seja mais rápida e realizada de maneira não destrutiva (Beć; Grabska; Huck, 2020)

A utilização da NIRS sempre foi negligenciada devido à baixa quantidade de informação que poderia ser obtida por meio dos espectros, na qual a região apresenta bandas de absorção amplas, com sobreposição e fracas. Entretanto na década de 80, houve um grande número de trabalhos publicados tornando-a uma técnica que até os dias atuais é amplamente utilizada pela indústria, especialmente na identificação de umidade em alimentos, e na agricultura, onde é amplamente utilizada nos dias atuais (Torniainen et al., 2020)

A espectroscopia do infravermelho próximo torna possível a realização de análises de maneira rápida, precisa, eficaz, e com baixo custo por não ser necessário o preparo da amostra. Essa técnica pode ser utilizada para a análise físico-química qualitativa e também até de maneira quantitativa, tornando esta técnica excelente para análise de rotina (EMBRAPA, 2014).

Devido ao grande número de informação que é obtida por meio dos espectros, a quimiometria é uma ferramenta que está diretamente associada a espectroscopia do infravermelho próximo. Dentre as principais técnicas a análise de componentes principais (PCA) é a principal, na qual é possível obter pesos e porcentagem de variância extraída de cada dimensão comum, associados as diferenças nas composições químicas das amostras analisadas (Machado et al., 2023)

3.5 QUIMIOMETRIA

As primeiras referências a quimiometria é datada da década de 70, como uma tradução de um termo sueco “kemometri”. Atualmente, a definição mais aceita para a quimiometria é que “a quimiometria é a disciplina da química que une a matemática, a estatística, e métodos utilizando a lógica formal. Visam projetar e selecionar experimentos e procedimentos ótimos de medição para fornecer o máximo de informação química relevante, por meio da análise dos dados químicos” (Aleixandre-Tudo et al., 2022)

Desse modo, a quimiometria pode ser definida como a utilização de métodos matemáticos e estatísticos utilizando-se de dados químicos, para análises qualitativas e quantitativas. A quimiometria, antes mesmo de ser definida na década de 70, já era utilizada, juntamente com a utilização de métodos estatísticos univariados, em dados químicos. Com o surgimento de computadores cada vez mais avançados, houve um grande aumento da utilização da quimiometria, intensificando assim o número de pesquisas na área, o que possibilitou o desenvolvimento de metodologias multivariadas para a interpretação de dados químicos, em áreas como a cromatografia e na espectroscopia (principalmente NIRS) (García Martín, 2022)

Os métodos quimiométricos podem ser subdivididos em dois principais, sendo eles os métodos exploratórios, sendo esses mais utilizados para a classificação ou discriminação, e métodos de predição, sendo esses métodos voltados a calibração e quantificação de compostos em diferentes tipos de amostras (Aleixandre-Tudo et al., 2022).

Grandes conjuntos de dados são amplamente geradas em diversas áreas, e visando interpretar as informações contidas nesses dados, diversos métodos foram desenvolvidos, a fim de conseguir reduzir a dimensionalidade desses dados, preservando assim as informações. A PCA é uma das mais antigas e mais utilizadas técnicas, tendo ela um objetivo simples, diminuir a dimensionalidade dos dados, preservando a variabilidade. Essa preservação da variabilidade consiste em encontrar novas variáveis que são funções lineares dos dados originais, maximizando a variância e diminuindo a correlação entre elas. Essa nova variável encontrada é denominada componente principal (Jolliffe; Cadima, 2016)

As componentes principais são encontradas de modo que a componente principal 1 (PC1) é a que apresenta a maior explicabilidade do conjunto de dados, e a

PC2 é a segunda que melhor explica, de modo que grandes quantidades de informações que inicialmente foram geradas, possam ser explicadas por um número reduzido de componentes principais (Ferreira, 2015)

A técnica de Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) tem se consolidado como uma das ferramentas quimiométricas mais eficazes para a classificação supervisionada de amostras em estudos envolvendo espectroscopia. Derivada do método PLS (*Partial Least Squares*), o PLS-DA permite a construção de modelos preditivos que correlacionam dados espectrais com classes previamente definidas. Sua aplicação tem sido amplamente explorada em áreas como controle de qualidade, autenticação de produtos e rastreabilidade na indústria alimentícia (Mehmood et al., 2012).

A integração de técnicas espectroscópicas avançadas — como NIR, FT-IR e Raman — com métodos quimiométricos como o PLS-DA tem revolucionado a análise de alimentos, oferecendo alternativas rápidas, não destrutivas e ambientalmente sustentáveis aos métodos convencionais. O estudo destaca que o PLS-DA é particularmente eficaz na discriminação de amostras com composição química semelhante, mas pertencentes a diferentes classes, como variedades de frutas, tipos de grãos ou espécies de café (Kharbach et al., 2023).

No contexto do café, diversos estudos recentes demonstram a eficácia do PLS-DA na diferenciação de espécies, origens geográficas e até métodos de processamento (Nguyen Minh et al., 2022) Os autores aplicaram PLS-DA a espectros NIR de 153 amostras de café verde vietnamita, com o objetivo de autenticar a espécie e a origem geográfica. Os modelos construídos com pré-processamento por derivada de primeira ordem e normalização variada apresentaram bom desempenho.

Outro estudo foi conduzido por SIM et al., 2024 que utilizaram PLS-DA para rastrear a origem de grãos de café especiais provenientes de três continentes, oito países e 22 regiões. Os espectros NIR foram processados com correção multiplicativa de espalhamento e centramento da média. O modelo PLS-DA apresentou alta acurácia na classificação por continente e país, demonstrando potencial para aplicações em sistemas de rastreabilidade. Embora modelos não lineares como Random Forest tenham superado o PLS-DA em acurácia regional, a técnica linear mostrou-se eficaz e interpretável para aplicações práticas.

Além disso, DHARMAWAN e MASITHOH (2025) aplicaram PLS-DA em conjunto com PCA e LDA para autenticar a origem de grãos de café robusta de quatro

regiões da Indonésia. Embora o modelo LDA combinado com redes neurais tenha alcançado a maior acurácia (100%), o PLS-DA também apresentou desempenho elevado, com acurácia superior a 96%, reforçando sua aplicabilidade em cenários industriais e laboratoriais.

Esses estudos evidenciam que o PLS-DA continua sendo uma ferramenta poderosa para análise de dados espectrais em amostras de café, especialmente quando combinado com estratégias de pré-processamento e seleção de variáveis. Sua capacidade de lidar com dados complexos e construir modelos interpretáveis o torna ideal para aplicações em alimentos, onde fatores como composição química, origem geográfica e variações sazonais influenciam diretamente os resultados analíticos.

Além da seleção de variáveis, entendida como a escolha de comprimentos de onda ou regiões espectrais mais relevantes para a discriminação entre as classes, o desempenho de modelos quimiométricos como o PLS-DA depende fortemente da qualidade dos dados espectrais utilizados. Espectros obtidos por NIR podem ser afetados por fatores físicos como espalhamento de luz, sensibilidade do detector, mudança no caminho óptico, variações na granulometria da amostra, densidade e umidade, que introduzem ruído e dificultam a extração de informações químicas relevantes. Para minimizar esses efeitos, técnicas de correção espectral como *Multiplicative Scatter Correction* (MSC) e *Standard Normal Variate* (SNV) são amplamente empregadas (Ferreira, 2015)

O MSC atua corrigindo os espectros com base em um espectro de referência, ajustando multiplicativamente e aditivamente cada amostra para reduzir o impacto do espalhamento. Já o SNV realiza uma normalização individual de cada espectro, centrando-o na média e escalando pela variância, o que é especialmente útil em amostras com variações físicas significativas. Ambas as técnicas têm se mostrado eficazes na melhoria da linearidade dos dados e na redução de interferências não químicas, contribuindo diretamente para o aumento da sensibilidade e exatidão dos modelos PLS-DA (Abrantes et al., 2023)

4 MATERIAIS E MÉTODOS

4.1 AMOSTRAS

As amostras de café das safras de 2020 e 2022 foram cedidas pelo Ministério da Agricultura e Pecuária (MAPA), conforme a chamada Nº 20/2018 Embrapa/ Consórcio Pesquisa Café, ConCafé, vinculado ao projeto “Soluções tecnológicas convergentes para a detecção de adulterantes, defeitos e espécie de café em grãos e café torrado e moído comercializados no Brasil”. As amostras utilizadas são provenientes dos seis principais estados produtores de café, sendo eles: Minas Gerais, Bahia, Espírito Santo, Paraná, Rondônia e São Paulo.

Foram utilizados um total de 154 amostras de café da safra de 2022, onde essas mesmas amostras foram moídas. Foram também utilizadas 53 amostras de café cru em grão e 21 amostras de café cru moído da safra de 2020. As amostras da safra de 2022 foram utilizadas para a calibração dos modelos e para a validação as amostras da safra de 2020.

As amostras de café em grão foram acondicionadas em ambiente com temperatura e umidade controlada. Para o processo de moagem, os grãos de café cru foram congelados com nitrogênio líquido, moídos em um moedor de café comercial e posteriormente peneirados (32 mesh) para homogeneização do tamanho de partícula.

4.2 OBTENÇÃO DOS ESPECTROS

Os espectros de infravermelho próximo foram obtidos utilizando o FOSS XDS *Rapid Content Analyzer* (Foss, Estados Unidos da América, Eden Prairie MN, USA). A faixa espectral foi de 400 a 2500 nm, com uma resolução de 2 nm, o espectro gerado é obtido pela média de 32 varreduras.

Os espectros foram obtidos em reflectância, de modo que a luz incidente penetra superficialmente nas amostras, sofrendo dispersão e parcialmente refletida de volta ao detector. A reflectância (R) obtida é transformada em absorbância por meio da relação da Equação 1.

$$A = \log\left(\frac{1}{R}\right) \quad Eq. (1)$$

Para o café moído, utilizou-se um “porta amostras esférico”, adequados para análise em pó, com aproximadamente 2,00 g de amostra, sendo que cada análise foi realizada em duplicata. Para o café em grão, foi utilizado um “porta amostras

retangular”, adequado para amostras em grãos, com aproximadamente 100,0 g de amostra.

4.3 ANÁLISE QUIMIOMÉTRICA E SOFTWARES

As análises quimiométricas foram realizadas utilizando o software MATLAB 2018®, por meio da interface gráfica GAMMA-GUI(Galvan et al., 2025). Os espectros obtidos foram submetidos a pré-processamentos para reduzir os efeitos físicos e melhorar a relação entre o sinal e a composição química das amostras.

Os pré-processamentos realizados foram os seguintes:

- Correção por Dispersão Multiplicativa (MSC), Variável Normal Padrão (SNV) e a primeira derivada (janela = 7, polinômio segundo grau), para minimizar os efeitos de espalhamento de luz e deslocamentos da linha de base.

Para a análise das amostras foram utilizadas as técnicas quimiométricas:

- Análise por Componentes Principais, para a visualização dos agrupamentos e identificação da variabilidade entre as amostras.
- Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) para a classificação das amostras de acordo com as diferenças dos espectros obtidos. Para a análise do PLS-DA os dados foram centrados na média, e o número de variáveis latentes foi determinado por meio da validação cruzada k-fold (k=5). O número ótimo de variáveis latentes escolhida foi a que apresentou a maior área sob a curva ROC (AUCCV)

Além disso foi desenvolvido um algoritmo para a seleção de variáveis para a discriminação das duas espécies. Para o desenvolvimento do algoritmo, foram fornecidos os espectros NIR dos cafés crus da safra de 2022 (moído e em grão), com os comprimentos de onda previamente selecionados. O algoritmo teve como objetivo encontrar razões espectrais entre os comprimentos de onda, de modo a maximizar a distância intergrupo (entre as duas espécies) e minimizar a distância intragrupo (dentro da mesma espécie).

O desempenho dos modelos gerados foi avaliado utilizando o coeficiente de correlação de Spearman (ρ), onde inicialmente as amostras são enumeradas em ordem crescente com base nas espécies (0 e 1), assim como os valores dos *ratios* obtidos nos modelos são classificados por ordem crescente e o coeficiente é obtido pela Equação 2.

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2-1)} \quad \text{Eq. (2)}$$

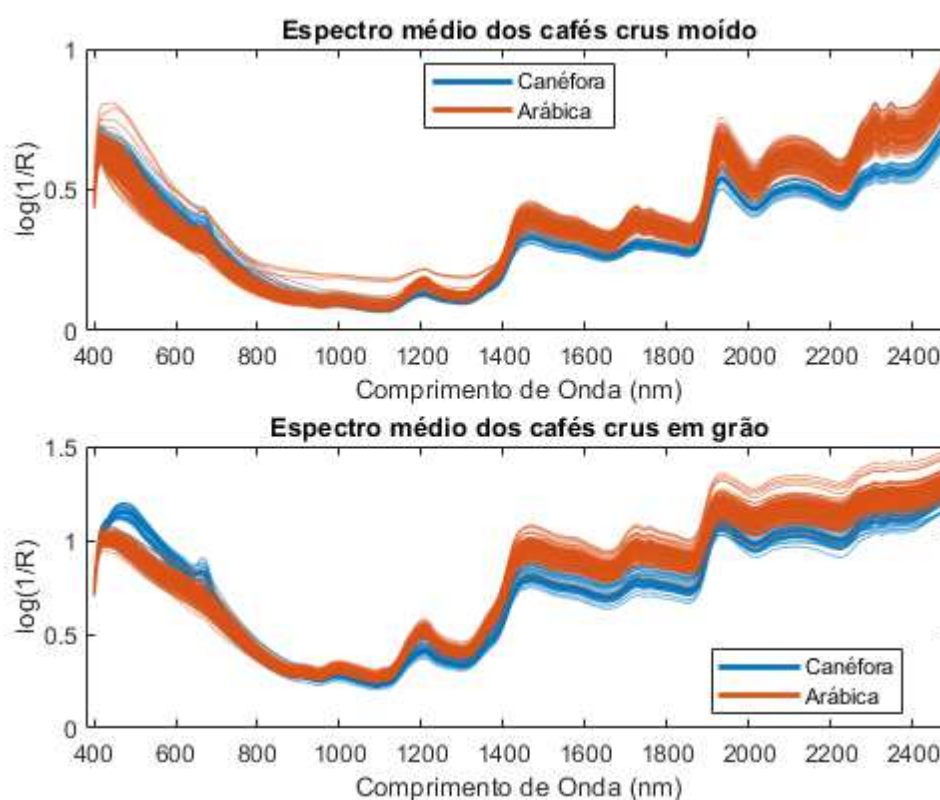
em que d_i é a diferença entre a classificação obtida com base no *ratio* do modelo e a classificação correspondente as amostras. Valores de ρ próximos de +1 indicam que o modelo classifica corretamente as amostras em relação às suas espécies, possuindo assim uma maior capacidade discriminante. Nesse processo iterativo, o modelo com maior correlação foi mantido, enquanto aquele com menor desempenho foi descartado (Ali; Al-Hameed, 2022).

5 RESULTADOS E DISCUSSÕES

5.1 ESPECTROS INFRAVERMELHO PRÓXIMO

Para verificar a diferença entre os espectros de infravermelho próximo das amostras das duas espécies distintas de café, foi obtido um espectro médio entre as duplicatas das amostras moídas e em grãos, cujos espectros médios dos dois tipos de cafés são apresentados na Figuras 3.

Figura 3 - Espectro NIRS das amostras de espécie de cafés crus arábica e canéfora moído e em grão



Fonte: o próprio autor

Em ambas as amostras de café, em grãos e moído, foi possível verificar o mesmo comportamento, em que a faixa do espectro de 400 a 700 nm apresentou um sinal de maior intensidade para os cafés canéforas, enquanto em comprimentos de onda de 900 a 2500 nm, os cafés arábica apresentam a maior intensidade.

Essa diferença da intensidade pode ser causada por alguns fatores, sendo como um dos principais a diferença de concentração de macros nutrientes, como carboidratos, proteínas e lipídeos, ou compostos como cafeína, trigonelina e ácidos clorogênicos. Além disso, pode ser causada por efeito de espalhamento de luz, sendo

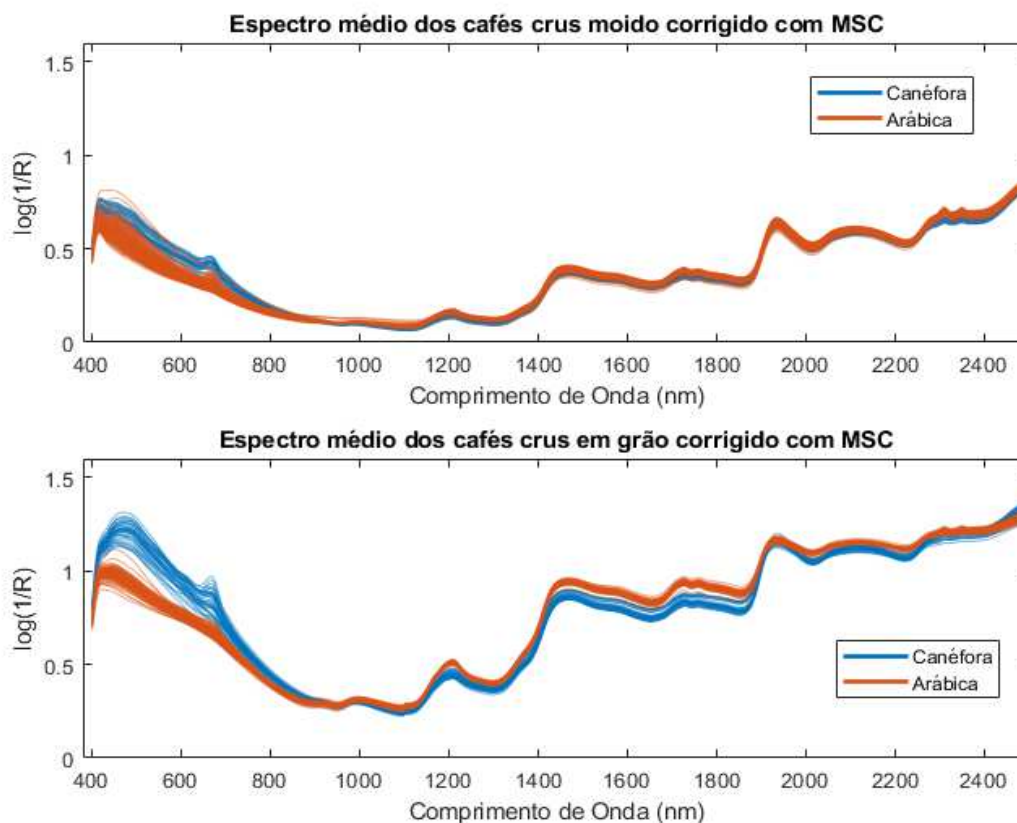
causada devido a variação do tamanho das amostras. Esse efeito, no entanto, pôde ser minimizado por meio de técnicas quimiométricas.

Entre as técnicas quimiométricas utilizadas para minimizar esses efeitos, o *Multiplicative Scatter Correction* (MSC), é uma das mais utilizadas para a correção dos efeitos relacionados ao espalhamento de luz em espectros NIRS. O MSC busca reduzir as variações não químicas, de modo que as diferenças das intensidades sejam apenas correlacionadas a composição das amostras.

Outras técnicas também são utilizadas para minimizar os efeitos de espalhamento e correção da linha de base, como a Variável Normal Padrão (SNV) e a primeira derivada. As principais diferenças entre os três métodos é que enquanto o MSC utiliza um espectro médio entre as amostras para realizar essa correção, o SNV e a primeira derivada com alisamento Savitzky Golay realizam essa correção com base apenas no espectro em si.

O espectro NIR corrigido com MSC é apresentado na Figura 4.

Figura 4 - Espectros NIRS das amostras de cafés arábica e canéfora crus em grãos e moído corrigido com MSC

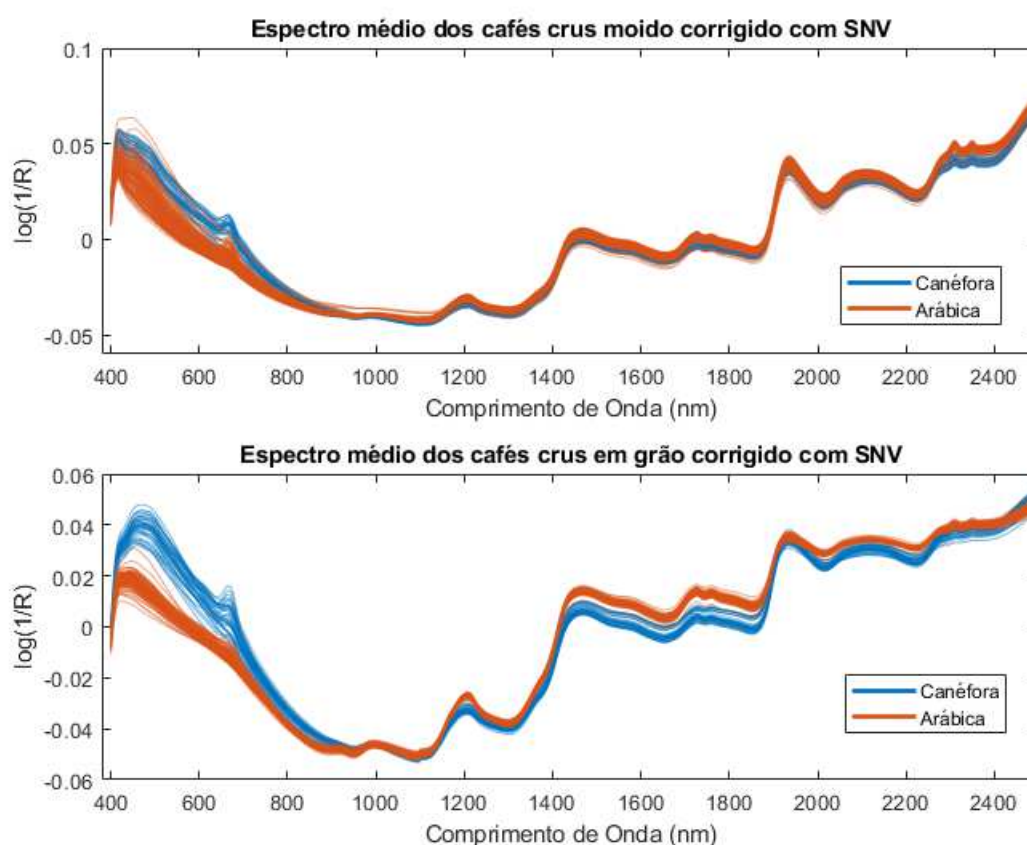


Fonte: o próprio autor

Com a aplicação do MSC no espectro bruto, tanto do moído quanto em grão, os espectros apresentam uma maior sobreposição, evidenciando assim o a presença de um efeito de espalhamento de luz, mantendo assim algumas regiões que apresentam diferença entre os dois grupos de amostra, principalmente de 2200 a 2400nm, e 1000 a 1400nm. Para o grão a região de 400 a 800 apresenta uma grande diferença entre as duas espécies, assim como diferenças em todo o espectro NIR (1000 a 2400 nm).

Na Figura 5 é apresentada os espectros corrigidos com o SNV.

Figura 5 - Espectros NIRS das amostras de cafés arábica e canéfora crus em grãos e moído corrigido com SNV

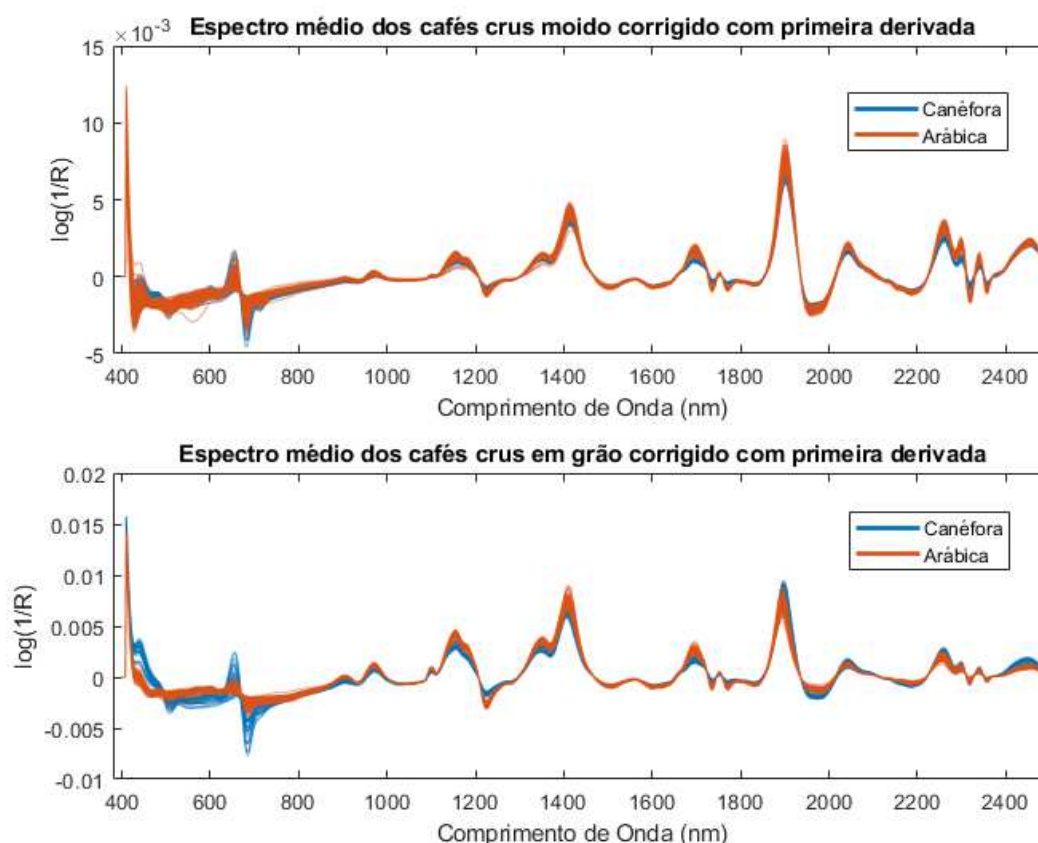


Fonte: o próprio autor

No espectro do café moído corrigido com o SNV, quando comparado ao MSC, observa-se uma menor sobreposição dos espectros em algumas regiões, como nas faixas de 1600 a 1900 nm e de 2200 a 2400 nm. Para o café em grão, ocorre o mesmo efeito, sendo mais evidente na região de 1600 a 1900 nm, especialmente para o grupo de cafés canéforas. Todavia, assim como no MSC, ocorre uma grande variação nas intensidades da região de 400 a 800 entre as duas espécies.

Além do MSC e do SNV, a primeira derivada com alisamento Savitzky Golay (janela = 7, polinômio de segundo grau) também foi realizada, o espectro corrigido é apresentado na Figura 6.

Figura 6 - Espectros NIRS das amostras de cafés arábica e canéfora crus em grãos e moído corrigido com primeira derivada com alisamento Savitzky Golay



Fonte: o próprio autor

No espectro do café moído corrigido com a primeira derivada com alisamento Savitzky Golay, observa-se que as curvas médias das duas espécies apresentam elevada sobreposição ao longo da maior parte do espectro, com diferenças mais evidentes em algumas regiões específicas, como nas faixas de aproximadamente 1100 a 1300 nm e em torno de 1600 a 1800 nm. Para o café em grão, comportamento semelhante é observado, sendo que as diferenças se tornam mais perceptíveis principalmente na região de 600 a 800 nm e em torno de 1600 a 1800 nm, especialmente para o grupo dos cafés canéforas.

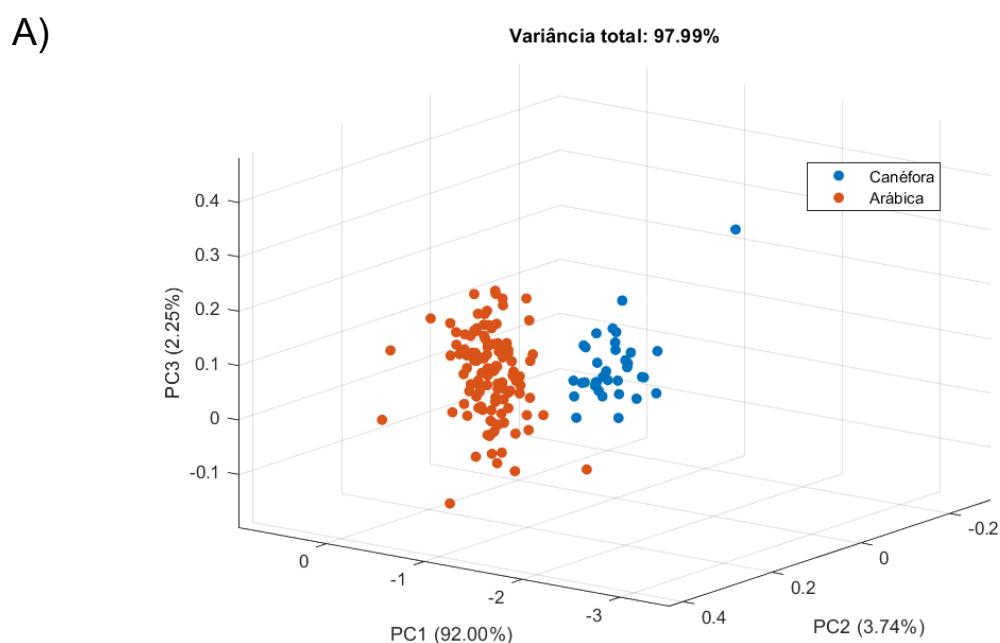
Devido à natureza da correção, os espectros corrigidos com o MSC preservam mais informação de intensidade absoluta, enquanto nos outros dois métodos, ocorrem uma variação maior da intensidade.

Após a aplicação dos pré-processamentos nos espectros do café cru, tanto moído quanto no em grão, observa-se uma melhor sobreposição entre as amostras, evidenciando a redução dos efeitos de espalhamento de luz, tornando os espectros mais consistentes e apropriados para as análises quimiométricas, como a PCA e PLS-DA, descritas nas sessões subsequentes.

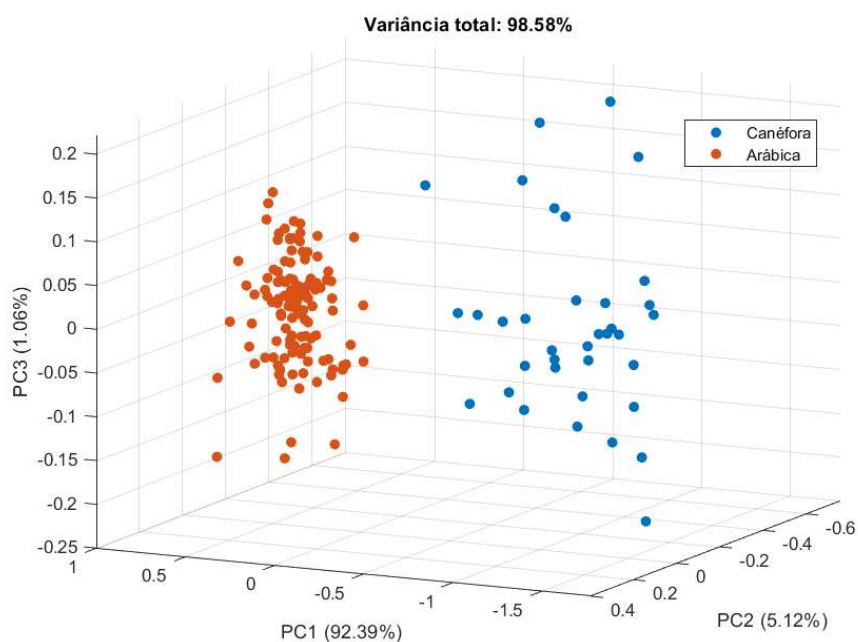
5.2 ANÁLISE POR COMPONENTES PRINCIPAIS (PCA)

A Análise por Componentes Principais é uma técnica exploratória que permitiu verificar a diferenciação entre as duas espécies de café. Sendo possível identificar por meio dos gráficos dos *scores* que tende a ocorrer a separação das duas espécies em grupos distintos, evidenciando assim a diferença espectral entre os dois grupos. Para a realização da análise por componente principal foram utilizados apenas as regiões do espectro NIR, assim, os comprimentos de onda de 400 a 1100 nm foram removidos, sendo utilizado apenas a faixa de 1100 a 2498 nm. Os scores das PCs 1, 2 e 3 para o café moído e em grão com os três pré-processamentos são apresentados nas figuras 7, 8 e 9.

Figura 7 - Gráfico de *scores* para as amostras de café cru-utilizando MSC, para o café moído (A) e em grão (B)



B)



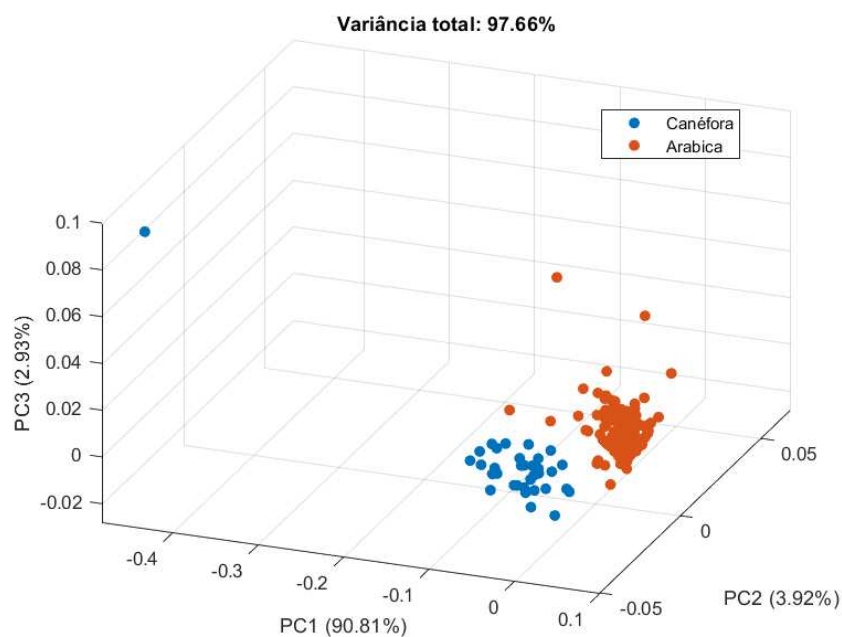
Fonte: o próprio autor

Com a representação tridimensional dos scores da PCA (PC1, PC2 e PC3) para o café moído, utilizando espectros corrigidos por MSC, observa-se uma separação clara entre as amostras de café canéfora e arábica. A distribuição dos pontos no espaço tridimensional evidencia a formação de dois grupos bem definidos, com baixa sobreposição entre as classes, indicando alta capacidade discriminatória do modelo. A variância total explicada pelas três primeiras componentes principais foi de 97,99%, sendo a PC1 responsável pela maior parte da variabilidade dos dados, enquanto PC2 e PC3 contribuem para o refinamento da separação entre os grupos.

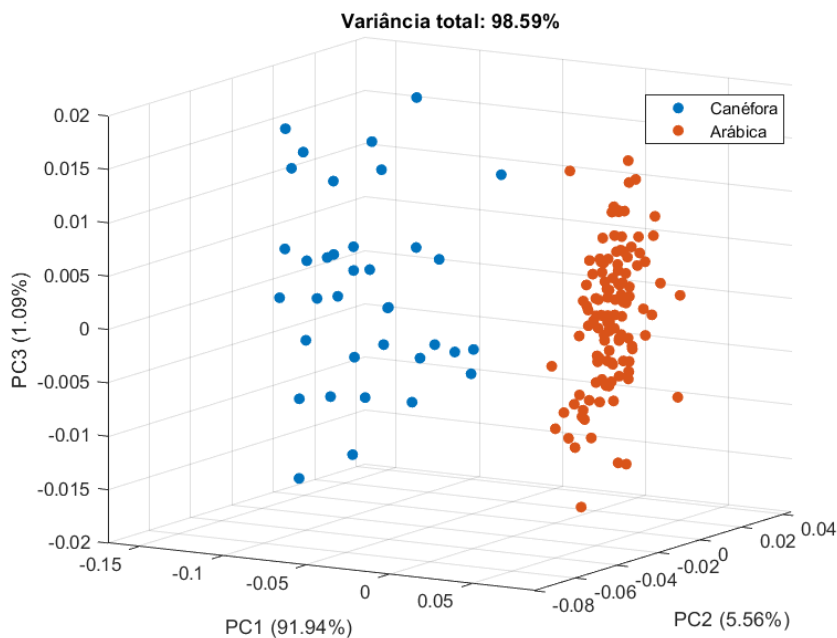
Para o café em grão, também se observa a formação de agrupamentos distintos correspondentes às espécies avaliadas, com separação evidente entre as classes. Nesse caso, a discriminação ocorre predominantemente ao longo da PC1, que concentra a maior parte da variância explicada (92,39%), enquanto as componentes PC2 e PC3 apresentam contribuições menores, atuando principalmente na dispersão dos dados dentro de cada grupo. A variância total explicada foi de 98,58%, indicando uma boa capacidade de representação do conjunto de dados.

Figura 8 - Gráfico de scores para as amostras de café cru-utilizando SNV, para o café moído (A) e em grão (B)

A)



B)



Fonte: o próprio autor

Com a representação tridimensional dos scores da PCA (PC1, PC2 e PC3) para os espectros corrigidos por SNV, observa-se uma elevada variância explicada para o café moído (97,66%) e para o café em grão (98,59%), indicando que as três primeiras componentes principais são capazes de representar adequadamente a estrutura dos

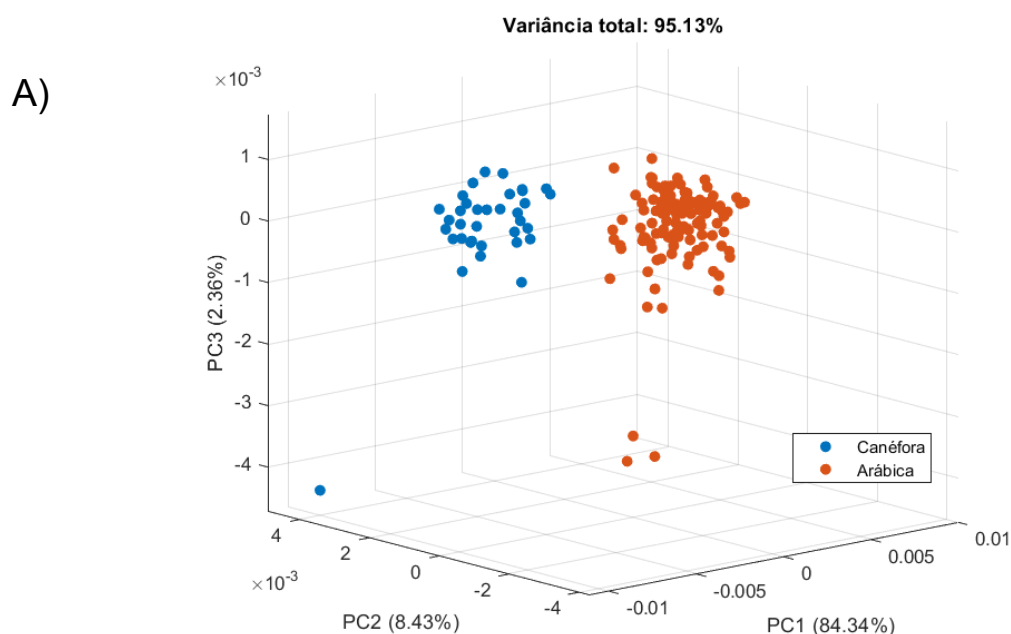
dados.

Para o café moído, a distribuição das amostras no espaço tridimensional evidencia a formação de dois grupos correspondentes às espécies avaliadas. A separação entre café arábica e canéfora ocorre de forma clara, sendo influenciada principalmente pela PC1, com contribuição complementar das componentes PC2 e PC3 na organização e definição dos agrupamentos.

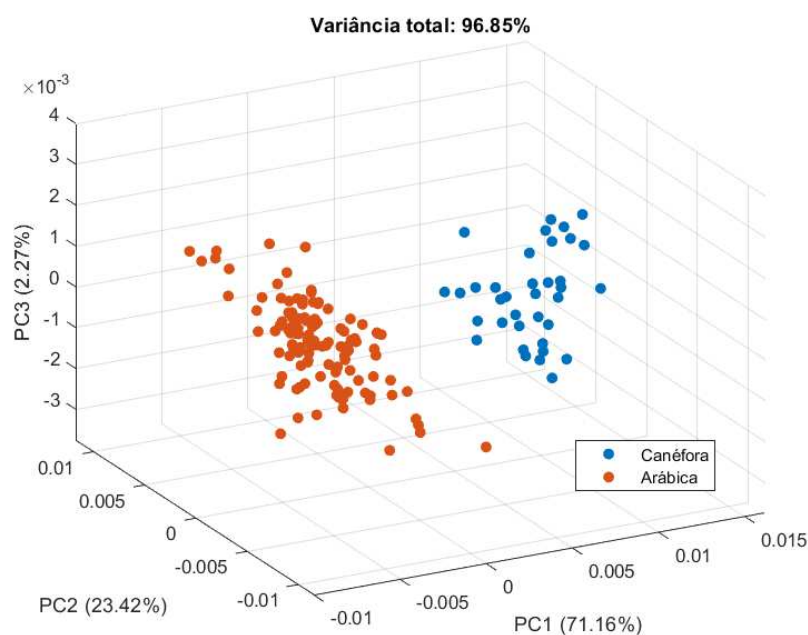
Para o café em grão, também se observa a formação de dois agrupamentos distintos, com separação evidente entre as espécies. Nesse caso, a PC1 apresenta papel predominante na discriminação entre os grupos, enquanto as componentes PC2 e PC3 contribuem principalmente para a dispersão das amostras dentro de cada classe.

De maneira geral, a análise tridimensional dos scores obtidos com a correção SNV confirma a capacidade da PCA em discriminar as espécies de café, apresentando comportamento semelhante ao observado para os dados corrigidos por MSC.

Figura 9 - Gráfico de *scores* para as amostras de café cru-utilizando alisamento Savitzky Golay e primeira derivada, para o café moído (A) e em grão (B)



B)



Fonte: o próprio autor

Com a utilização da primeira derivada nos espectros, a análise por PCA em três dimensões (PC1, PC2 e PC3) também permitiu a discriminação entre as amostras de café arábica e canéfora. A variância total explicada pelas três primeiras componentes principais foi de 95,13% para o café moído e 96,85% para o café em grão, indicando boa capacidade de representação dos dados.

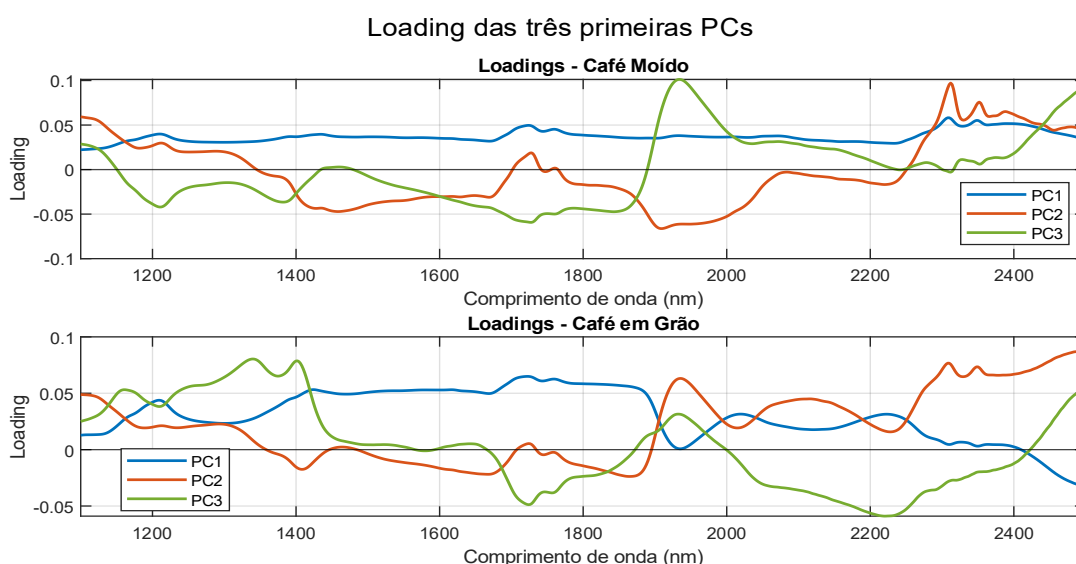
Para o café moído, observa-se a formação de agrupamentos associados às espécies, porém com menor definição quando comparado aos pré-processamentos anteriores. A separação entre arábica e canéfora é menos evidente, sugerindo maior sobreposição entre as classes no espaço tridimensional. Nesse caso, embora a PC1 concentre a maior parte da variância e seja a principal responsável pela discriminação, a PC2 também contribui para a separação entre os grupos, ainda que de forma menos pronunciada, enquanto a PC3 apresenta contribuição mais limitada na distinção das classes.

Para o café em grão, por outro lado, a formação de dois agrupamentos distintos é mais evidente, indicando melhor desempenho da primeira derivada para esse tipo de amostra. A separação ocorre predominantemente ao longo da PC1, enquanto as componentes PC2 e PC3 auxiliam na organização da dispersão dos dados dentro de cada grupo.

Com os gráficos de Scores e as variâncias explicadas para os três pré-processamento, foi possível verificar que ambos conseguiram separar as duas espécies em dois grupos compostos sendo um majoritariamente arábica (A) e outro majoritariamente canéfora (C), onde o MSC foi o pré-processamento que apresentou a maior variância explicada. Com isso é possível verificar por meio do gráfico de *loadings* quais regiões do espectro influenciam essa separação, de modo a realizar uma seleção de variáveis para maximizar essa distância.

Os gráficos de *loadings* das três componentes principais são apresentados na Figura 10.

Figura 10 - Gráfico de *loadings* da PC1, PC2 e PC3 para as amostras de café cru moído e em grão



Fonte: o próprio autor

Os gráficos de *loadings* da PCA permitem avaliar quais regiões espectrais mais contribuem para a variabilidade dos dados. Para o café moído, a PC1 apresenta valores relativamente baixos e distribuídos ao longo de todo o espectro, sem a presença de picos acentuados, indicando que a variabilidade capturada por essa componente está associada a contribuições amplas e distribuídas, e não a regiões espectrais específicas. Esse comportamento sugere que a separação entre as espécies ocorre devido a pequenas diferenças acumuladas ao longo de diversos comprimentos de onda.

Para o café em grão, os *loadings* da PC1 apresentam maior variação ao longo do espectro, com mudanças de sinal e regiões com maior magnitude, especialmente

em torno de determinadas faixas espectrais. Esse comportamento indica que, diferentemente do café moído, a variabilidade capturada pela PC1 está mais associada a regiões específicas do espectro.

Em relação à PC2, observa-se, para ambas as matrizes, a presença de regiões com variações mais pronunciadas nos *loadings*, com picos positivos e negativos mais definidos. Esse comportamento indica que a PC2 está associada a diferenças espectrais mais localizadas, contribuindo para a separação entre as espécies ao destacar variações específicas que não são capturadas pela PC1.

A PC3 apresenta variações ainda mais acentuadas e localizadas, com mudanças frequentes de sinal ao longo do espectro. Esse padrão sugere que essa componente está relacionada a variações mais sutis e específicas entre as amostras, podendo estar associada a diferenças de menor intensidade na composição química ou a efeitos residuais não explicados pelas componentes anteriores.

De modo geral, observa-se que a PC1 descreve a variabilidade global dos espectros, enquanto as componentes subsequentes (PC2 e PC3) contribuem para a discriminação entre as espécies ao evidenciar diferenças espectrais mais específicas e localizadas.

Todavia, por ser um modelo exploratório, a análise por componentes principais não permite a classificação das amostras em um dos dois grupos sendo, portanto, necessário a aplicação de um modelo discriminante.

5.3 ANÁLISE DISCRIMINANTE POR MÍNIMOS QUADRADOS PARCIAIS

A Análise Discriminante por Mínimos Quadrados Parciais, diferente da Análise por Componentes Principais é uma técnica quimiométrica supervisionada, na qual é modelada uma relação linear entre a matriz de variáveis independentes (X), representada pelos espectros, e a variável dependente (Y), correspondente às classes das amostras. Essa relação é obtida por meio da construção de variáveis latentes que maximizam a covariância entre X e Y, permitindo a extração de informações relevantes para a discriminação entre as classes.

A partir do modelo construído, obtém-se uma resposta contínua para cada amostra, a qual pode ser convertida em uma classificação por meio da definição de um limiar de decisão. Para o desenvolvimento e avaliação do modelo, são utilizados conjuntos de calibração e validação, ambos contendo amostras de todas as classes,

sendo o primeiro empregado na construção do modelo e o segundo na avaliação de sua capacidade preditiva.

Uma etapa importante para o bom desempenho do modelo é a seleção de variáveis, realizada por meio da atribuição de regiões do espectro NIR com ligações químicas presentes em compostos químicos, que podem ser encontrados em diferentes concentrações em ambas espécies de café, e que alguns deles podem variar de acordo com aspectos que não necessariamente estão somente correlacionados com a espécie, como maior ou menor precipitação de chuvas, e, maior ou menor altitude.

Entretanto, três compostos apresentam diferenças correlacionadas as espécies, a cafeína, a trigonelina e o ácido clorogênico. O ácido clorogênico e a cafeína estão presentes em maiores concentrações no *Coffea canephora*, enquanto a trigonelina está presente em maiores concentrações no *Coffea arabica*. Desse modo é possível correlacionar esses analitos com determinadas regiões do espectro NIR de modo a conseguir obter um modelo que não possa ser utilizado em amostras plantadas em diferentes regiões, assim como de safras diferentes (Adnan et al., 2020; Ky et al., 2001).

Foi realizado uma seleção de comprimentos de onda que possuem relação com os analitos presentes em concentrações distintas nas duas espécies de café, conforme apresentado na Tabela 2

Tabela 2 - Bandas, atribuições e compostos relacionado aos comprimentos de onda das bandas selecionadas

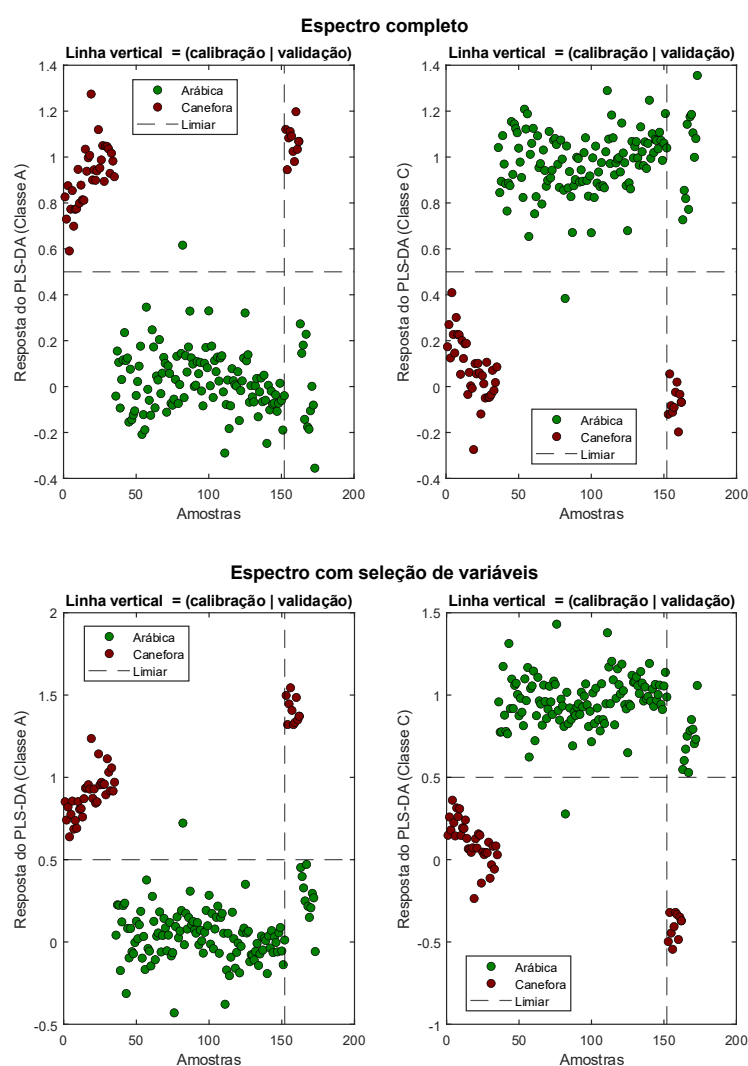
Bandas (nm)	Atribuições	Compostos relacionados
1140 - 1292	Segundo <i>overtone</i> do estiramento C-H	Cafeína e trigonelina
1358 - 1656	Primeiro <i>overtone</i> N-H e O-H	Ácidos clorogênicos e trigonelina
1664 - 1852	Primeiro <i>overtone</i> do estiramento C-H	Cafeína e trigonelina
1878 - 2016	Bandas de combinação C=O e O-H	Ácidos clorogênicos e cafeína
2016 – 2226	Bandas de combinação N-H	Trigonelina
2234 - 2400	Bandas de combinação C-H * CC (-CH, CH ₂ , CH ₃)	Cafeína e trigonelina

Fonte: (Ribeiro; Ferreira; Salva, 2011)

Com os espectros obtidos, com e sem a seleção de variáveis, foi realizada a

análise PLS-DA, com o objetivo de verificar se a seleção de variáveis contribui para a melhora da discriminação das espécies, especialmente na predição de amostras de café cru provenientes de outras safras. Para a construção do modelo PLS-DA foram utilizadas 3 variáveis latentes. Os resultados do PLS-DA estão apresentados na Figura 11.

Figura 11 - Respostas da PLS-DA para as amostras de café cru moído utilizando o espectro completo e com variáveis selecionadas



Fonte: o próprio autor

Para avaliar se a seleção das variáveis afetou positivamente a classificação das amostras, foram analisadas: a sensibilidade, a seletividade e a exatidão, podendo assim identificar se houve um aumento no número de amostras corretamente classificadas e, conseqüentemente redução de falsos positivos e falsos negativos.

Tabela 3 - Parâmetros de desempenho do PLS-DA para o café arábica cru moído

Parâmetros	Espectro completo		Com seleção de variáveis	
	Calibração	Validação	Calibração	Validação
Sensibilidade (%)	99,14	100	99,14	100
Seletividade (%)	100	100	100	100
Exatidão (%)	99,34	100	99,34	100

Fonte: o próprio autor

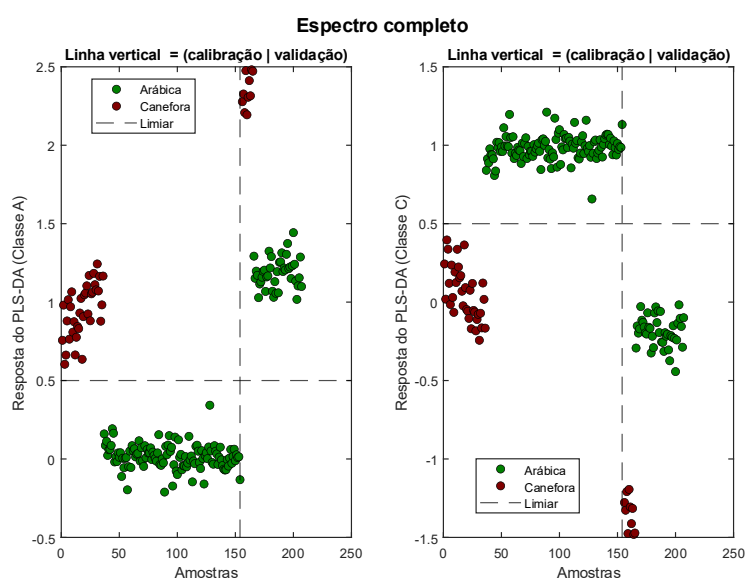
Tabela 4 – Matriz de confusão do PLS-DA para café moído classe arábica

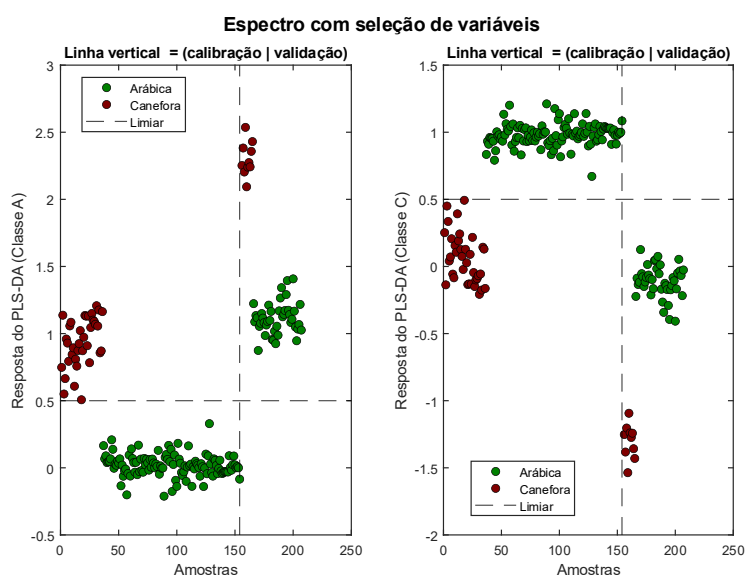
	Calibração		Validação	
	Positivo	Negativo	Positivo	Negativo
Positivo	116	1	11	0
Negativo	0	35	0	10

Fonte: o próprio autor

Por meio dos parâmetros avaliados para o modelo, é possível identificar que com a seleção de variáveis não houve um aumento performance da calibração, assim como continuou classificando corretamente todas as amostras do conjunto de validação.

Para as amostras de café em grão também foi realizada a análise PLS-DA e o resultado é apresentado na Figura 12

Figura 12 - Respostas da PLS-DA para as amostras de café cru em grão utilizando o espectro completo e com seleção de variáveis



Fonte: o próprio autor

Para ambos os modelos, o melhor resultado foi obtido utilizando-se duas variáveis latentes. A eficiência dos modelos foi avaliada por meio da sensibilidade, seletividade e exatidão, apresentados na Tabela 5.

Tabela 5 - Parâmetros de desempenho do PLS-DA para o café arábica cru em grão

Parâmetros	Espectro completo		Com seleção de variáveis	
	Calibração	Validação	Calibração	Validação
Sensibilidade (%)	100	0,00	100	0,00
Seletividade (%)	100	100	100	100
Exatidão (%)	100	18,87	100	18,87

Fonte: o próprio autor

Tabela 6 – Matriz de confusão do PLS-DA para café moído classe arábica

	Calibração		Validação	
	Positivo	Negativo	Positivo	Negativo
Positivo	118	0	0	43
Negativo	0	36	0	10

Fonte: o próprio autor

No conjunto de calibração o modelo se mostrou eficiente na classificação das amostras de café arábica, assim como para os cafés da espécie canephora. Para o conjunto de validação, as amostras de café canephora foram classificadas corretamente, enquanto todas as amostras dos arábicas foram classificadas erradas.

Este comportamento evidencia os efeitos das variações na composição dos grãos entre diferentes safras, os quais impactam diretamente os espectros obtidos por

espectroscopia no infravermelho próximo (NIR). A seleção de variáveis realizada está associada às ligações químicas presentes nos analitos que exibem diferenças de concentração entre as duas espécies analisadas. No entanto, tais ligações não são exclusivas desses compostos, podendo também estar presentes em macronutrientes como carboidratos, proteínas e lipídeos. Assim, embora os efeitos das variações sazonais tenham sido parcialmente minimizados, tais flutuações ainda podem comprometer a aplicabilidade do modelo.

Dessa forma, observa-se que o pré-processamento por MSC não apresentou resultados satisfatórios para a discriminação das espécies no contexto avaliado. Nesse sentido, o desenvolvimento do algoritmo baseado em razões espectrais surge como uma alternativa para a diferenciação entre as espécies de café, apresentando maior robustez frente às variações observadas entre as safras.

Os resultados para o PLS-DA para os espectros pré-processados com SNV e primeira derivada, assim como os parâmetros de desempenhos dos modelos são apresentados no Apêndice A.

5.4 DISCRIMINAÇÃO DA ESPÉCIE DE CAFÉ MOÍDO E EM GRÃO

Com base nos modelos PLS-DA foi possível identificar que a seleção de variáveis não impactou na efetividade dos modelos tanto para o café moído quanto para o café em grão. De modo a melhorar a seleção das variáveis para a separação dos grupos, foi utilizado um algoritmo desenvolvido em Python em parceria com a Embrapa Instrumentação (São Carlos, SP).

Após um período de 30 dias, o algoritmo conseguiu encontrar a melhor razão entre os comprimentos de onda que melhor separou os dois grupos de amostras, sendo assim encontrando uma Equação para os cafés moídos e uma outra Equação para os cafés em grãos. Esses comprimentos de onda estão apresentados nas Tabelas 7 e 8.

Tabela 7 - Comprimentos de onda do numerador e denominador para a separação das duas espécies em café moído

	Comprimento de onda
Numerador	890 – 892, 962 – 968, 998 – 1002, 1212, 1216, 1228 – 1230, 1356 – 1366, 1390, 1394, 1400, 1436 – 1442, 1446, 1568 – 1574, 1618 – 1620, 1624 – 1626, 1696 – 1700, 1896 – 1906, 1946 – 1956, 2076, 2086 – 2096, 2136 – 2146, 2166 – 2168, 2172 – 2174, 2304 – 2308, 2376 – 2386
Denominador	930 – 932, 998 – 1002, 1148 – 1150, 1194, 1198, 1212, 1216, 1436 – 1442, 1446, 1568 – 1574, 1618 – 1620, 1696 – 1700, 1722, 1758 – 1762, 1946 – 1956, 2066 – 2072, 2076, 2166 – 2168, 2304 – 2308, 2344 – 2348

Fonte: o próprio autor

Tabela 8 - Comprimentos de onda do numerador e denominador para a separação das duas espécies em café cru em grão

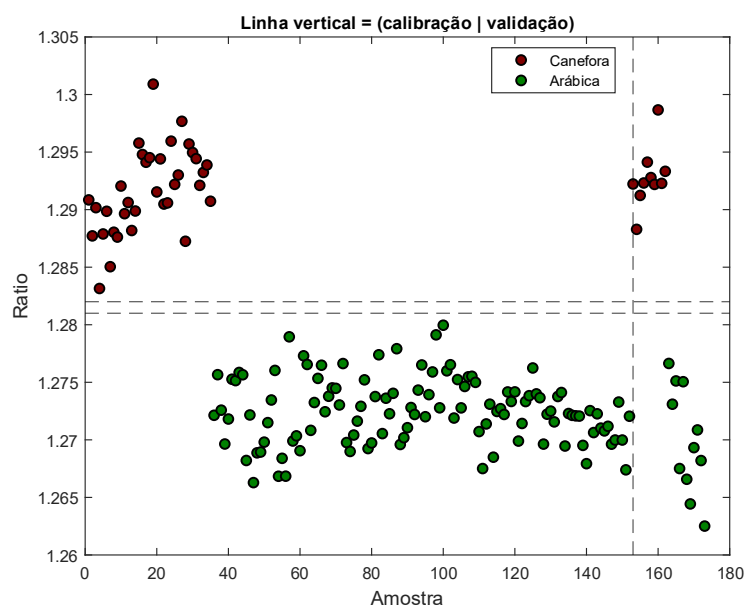
	Comprimento de onda
Numerador	890 – 892, 962 – 968, 998 – 1002, 1212, 1216, 1228 – 1230, 1356 – 1366, 1390, 1394, 1400, 1436 – 1442, 1568 – 1574, 1618 – 1620, 1624 – 1626, 1696 – 1700, 1896 – 1906, 1946 – 1956, 2076, 2086 – 2096, 2136 – 2146, 2166 – 2168, 2172 – 2174, 2304 – 2308, 2376 – 2386
Denominador	930 – 932, 998 – 1002, 1148 – 1150, 1194, 1198, 1212, 1216, 1436 – 1442, 1446, 1568 – 1574, 1618 – 1620, 1722, 1758 – 1662, 1946 – 1956, 2066 – 2072, 2076, 2166 – 2168, 2304 – 2308, 2344 – 2348

Fonte: o próprio autor

Com a Equação obtida pelo algoritmo que melhor separa as duas espécies, essa foi utilizada para a validação do modelo amostras da safra de 2020. Desse modo é possível utilizar essas amostras para verificar se efetivamente as Equações conseguem separar as duas espécies em amostras, que não estavam presente no conjunto utilizado na obtenção da Equação, para verificar se não houve um *overfitting*. Neste caso o modelo consegue separar apenas o conjunto de calibração, e quando aplicado a amostras que não pertencem a esse conjunto, a equação não é efetiva.

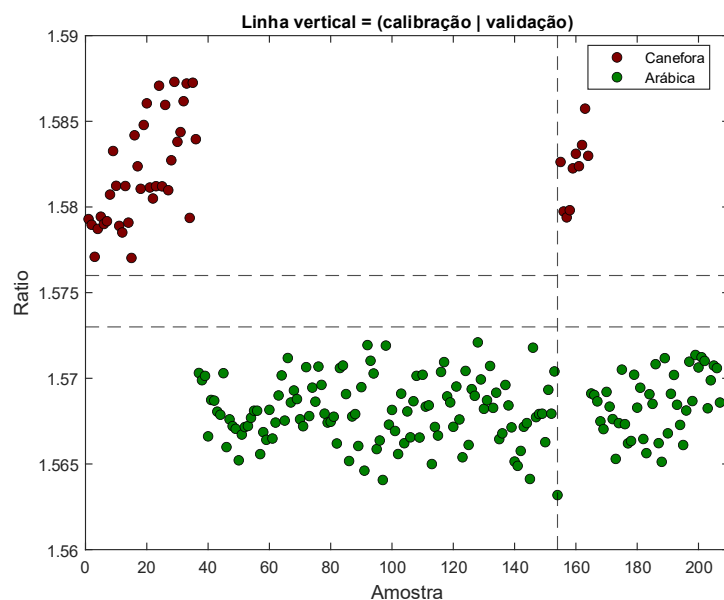
Os resultados dos *ratios* para as amostras de calibração e validação dos cafés crus moído e em grãos crus são apresentados na Figura 13 e 14.

Figura 13 – *Ratio* para as amostras de calibração e validação das amostras de café cru moído



Fonte: o próprio autor

Figura 14 - *Ratio* para as amostras de calibração e validação das amostras de café cru em grão



Fonte: o próprio autor

Para o modelo do café cru moído, a equação obtida pelo algoritmo se demonstrou efetiva na separação, assim evidenciando que com o modelo obtido é possível realizar a diferenciação das amostras de duas espécies de café de safras diferentes. Contudo, a distância entre os dois grupos é pequena, de modo que variam aproximadamente

0,002 unidades.

O modelo para a identificação da espécie nos grãos, assim como os do moído também de mostrou efetivo na separação das amostras de outra safra. Os dois grupos de amostras ficaram separados por uma diferença de 0,003 unidades. Desse modo, ambos os modelos se demonstraram efetivos para a separação das espécies de café cru.

Entretanto, analisando as Equações do ponto de vista químico, a utilização de comprimentos de ondas isolados em vez de bandas, ou parte delas, não apresentam sentido químico. Com isso, com o objetivo de dar um maior sentido químico, as Equações foram modificadas de modo a remover comprimento de ondas isolados, substituindo por bandas ou parte de bandas.

Assim como a soma dos comprimentos de onda foram substituídos pela integral das regiões selecionadas, desse modo, o efeito da resolução do equipamento é minimizada, a integral independe do número de pontos obtidos dentro dessa região, como demonstrado na equação abaixo:

$$Ratio = \frac{\sum_i \int_{b_i}^{a_i} B(\lambda) d\lambda}{\sum_j \int_{d_j}^{c_j} B(\lambda) d\lambda} \quad \text{Eq. (3)}$$

As faixas de integração do numerador e do denominador são apresentadas na Tabela 9 e 10

Tabela 9 - Comprimento de onda do numerador e denominador da razão para o café cru moído da Equação modificada

	Comprimento de onda (nm)
Numerador	1180 – 1218, 1392 – 1400, 1568 – 1574, 1946 – 1956, 2066 – 2072, 2136 – 2174
Denominador	1180 – 1216, 1436 – 1450, 1690 – 1700, 2066 – 2096, 2376 – 2386

Fonte: o próprio autor

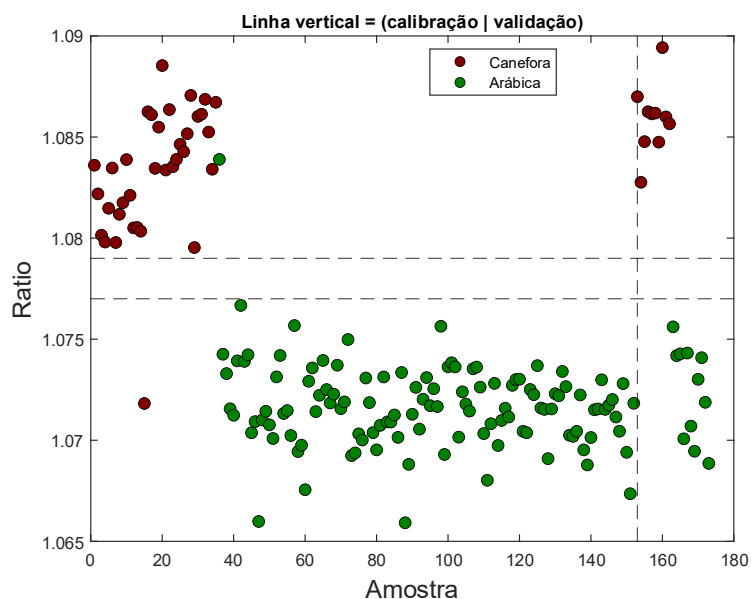
Tabela 10 - Comprimento de onda do numerador e denominador da razão para o café cru em grão da Equação modificada

	Comprimento de onda (nm)
Numerador	1180 – 1230, 1356 – 1366, 1390 – 1400, 1436 – 1446, 1564 – 1574, 1618 – 1626, 1696 – 1700, 1896 – 1906, 1946 – 1956, 2076 – 2096, 2166 – 2174, 2304 – 2308, 2376 – 2386
Denominador	1148 – 1160, 1194 – 1198, 1212 – 1228, 1436 – 1446, 1568 – 1574, 1696 – 1720, 1758 – 1762, 1946 – 1956, 2066 – 2076, 2166 – 2168, 2304 – 2308, 2344 – 2348

Fonte: o próprio autor

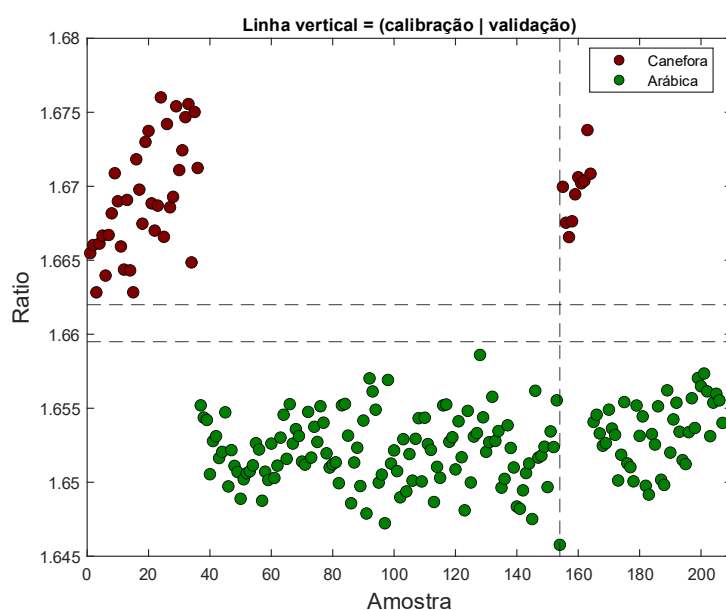
As novas Equações assim como as do algoritmo, foram obtidas utilizando o grupo de amostras de calibração, e posteriormente validadas com as amostras do grupo de validação. Os resultados dos *ratios* para as amostras de café cru moído e em grão são apresentados na Figura 15 e 16.

Figura 15 - Ratio para as amostras de calibração e validação das amostras de café cru moído



Fonte: o próprio autor

Figura 16 - Ratio para as amostras de calibração e validação das amostras de café cru em grão



Fonte: o próprio autor

Os *ratios* obtidos para as amostras de café cru moído utilizando a Equação modificada possibilitam identificar uma diminuição no número de acertos das amostras de calibração, onde uma amostra de cada espécie foi classificada de maneira errada. O aumento da explicabilidade química com a remoção de comprimentos de ondas isolados, substituídos por bandas do espectro faz com que ocorra a diminuição da exatidão do modelo. Quando realizado a validação com as amostras de outra safra, a classificação das amostras condiz com as espécies das amostras.

Tabela 11 - Parâmetros de desempenho do modelo baseado em uma Equação obtida por algoritmo e Equação modificada para café cru moído

Parâmetros	Equação do algoritmo		Equação modificada	
	Calibração	Validação	Calibração	Validação
Sensibilidade (%)	100	100	97,14	100
Seletividade (%)	100	100	95,15	100
Exatidão (%)	100	100	98,70	100

Fonte: o próprio autor

A distância intergrupo do modelo modificado, quando comparado com o modelo obtido por meio do algoritmo se mantém igual, com uma separação de 0,002 unidades, contudo, a distância intragrupo aumenta, tornando as amostras da mesma espécie mais distantes uma das outras.

Para as amostras de café em grão a Equação modificada, apresentou um desempenho de classificação equivalente ao obtido com a Equação gerada pelo algoritmo. No entanto, a adição de novos comprimentos de onda, aumentando a explicabilidade química, resultou em um aumento na distância intragrupo. A separação entre os grupos não foi alterada, mantendo a distância de 0,003 unidades entre as duas espécies.

Tabela 12 - Parâmetros de desempenho do modelo baseado em uma Equação obtida por algoritmo e Equação modificada para café cru em grão

Parâmetros	Equação do algoritmo		Equação modificada	
	Calibração	Validação	Calibração	Validação
Sensibilidade (%)	100	100	100	100
Seletividade (%)	100	100	100	100
Exatidão (%)	100	100	100	100

Fonte: o próprio autor

De modo geral, a Equação obtida por meio de algoritmo é a que maximiza a distância entre os dois grupos e a que melhor consegue diminuir a distância entre as amostras do mesmo grupo, sendo essa a melhor Equação matematicamente.

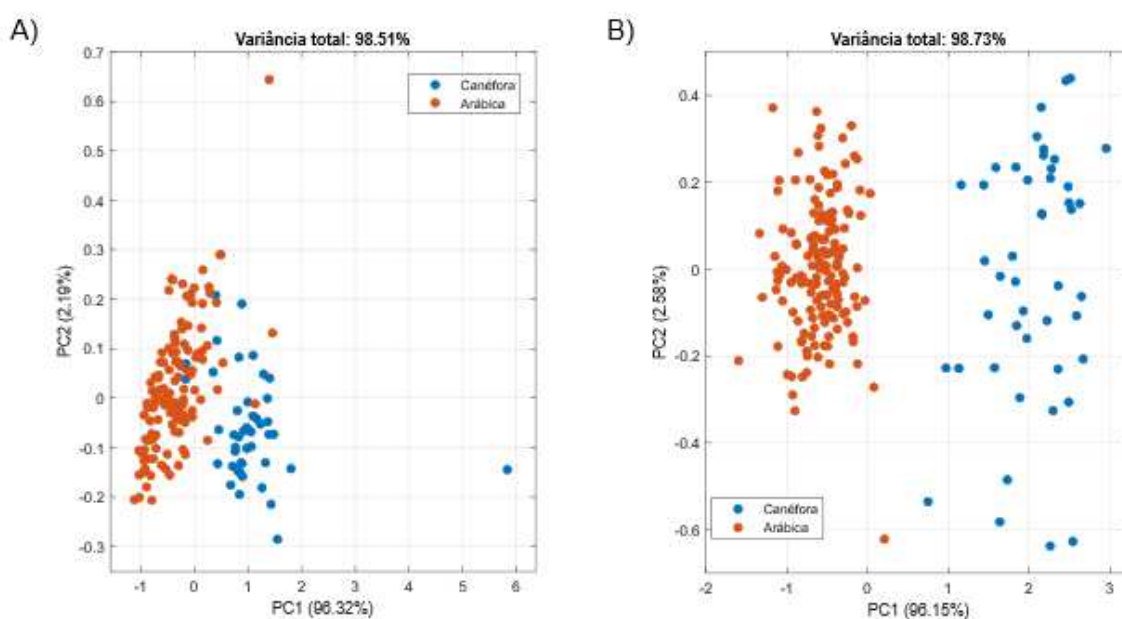
Entretanto, não possui muito sentido químico, visto que utiliza comprimentos de ondas isolados. A modificação da Equação mesmo diminuindo a distância entre os dois grupos, e aumentando entre as amostras do mesmo grupo, apresenta uma maior explicabilidade química, pois utiliza bandas definidas, assim como a utilização de integral para a obtenção dos *ratios*, possibilita a utilização da Equação em outros instrumentos com diferentes resoluções.

5.5 APLICAÇÃO DO ESPECTRO VISÍVEL NA DISCRIMINAÇÃO DAS ESPÉCIES DE CAFÉ

Os espectros foram obtidos em um equipamento VIS-NIR, e com o pré-processamento dos dados, é possível identificar uma diferença entre as duas espécies de café na região do visível (400 a 800 nm), embora a coloração dos grão não esteja diretamente correlacionado a espécie, ela está relacionada a presença de defeitos nos grão, e esses defeitos estão diretamente correlacionados a qualidade do produto.

Desse modo, foi realizado uma análise exploratória dos dados com a PCA, para verificar se os dois grupos podem ser diferenciados, e posteriormente uma análise discriminante, para efetivamente confirmar essa separação.

Figura 17 - Gráfico de Scores da PCA utilizando o espectro visível para o café cru moído (A) e em grão (B).

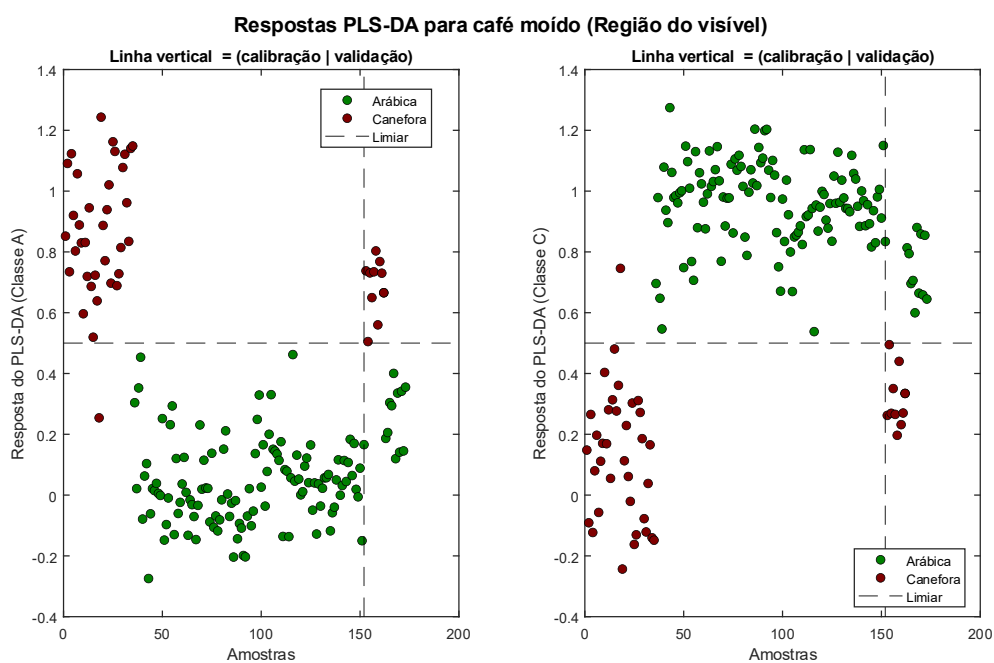


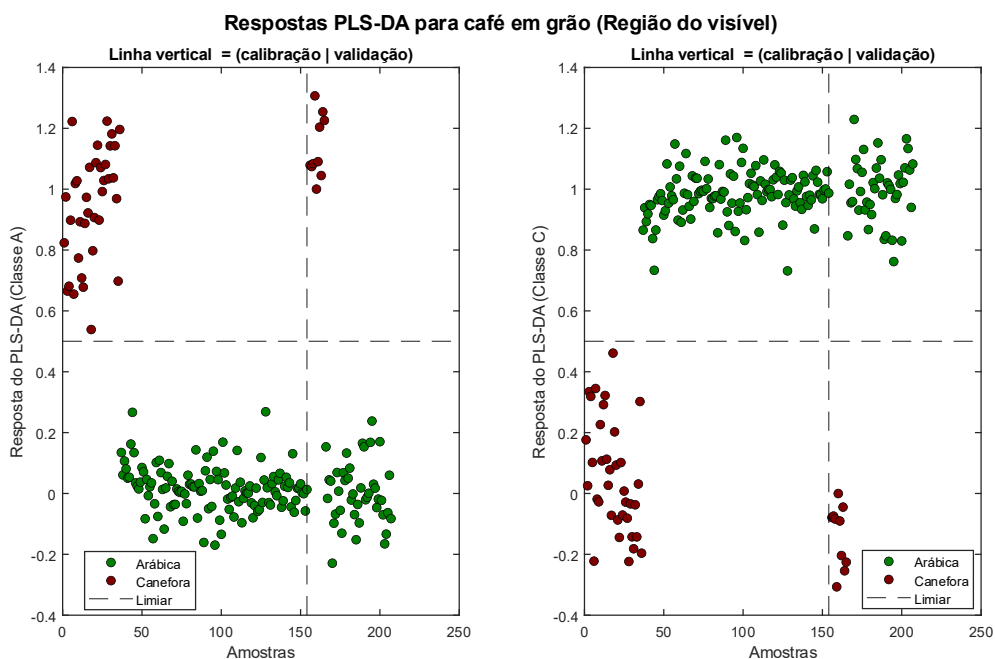
Fonte: o próprio autor

Com os Scores da PCA é possível identificar, para o café moído, que não existe a formação de dois grupos bem definidos. Em contrapartida, para os cafés em grão a formação de dois grupos é evidente, um grupo composto pelos cafés canéforas e outro pelos arábicas, evidenciando assim que existe uma diferença espectral entre as duas espécies nessa região.

Para efetivamente verificar se a separação ocorre, a análise por PLS-DA foi realizada, com a calibração feita com amostras da safra de 2022 e a validação com amostras da safra de 2020. Os resultados da análise são apresentados na Figura 18.

Figura 18 - Respostas da PLS-DA para as amostras de café cru moído e em grão utilizando o espectro visível





Fonte: o próprio autor

Utilizando o espectro da região do visível no modelo PLS-DA, tanto para o café moído quanto para o café em grão, é possível realizar a discriminação da espécie. Todavia, para garantir a efetividade dessa abordagem, é necessário o processo de beneficiamento do café, onde os defeitos e as impurezas são removidos. Uma vez que esses defeitos e impurezas podem não representar adequadamente a coloração da amostra, considerando que a coloração de grãos defeituosos não é necessariamente a mesma do café saudável.

6 CONCLUSÃO

Neste trabalho, foi desenvolvida uma metodologia baseada na espectroscopia no infravermelho próximo (NIRS) associada a métodos quimiométricos, com o objetivo de discriminar as espécies *Coffea arabica* e *Coffea canephora*, tanto em grão quanto moídas.

A PCA indicou que as duas primeiras componentes principais concentraram a maior parte da variância dos dados, e os gráficos de scores evidenciaram uma tendência de separação entre as espécies, devido a existência de diferenças espectrais. Além disso, os gráficos de *loadings* indicaram que essas diferenças estão distribuídas ao longo de grande parte do espectro, sem a predominância de uma única região espectral, o que reforça a necessidade de abordagens multivariadas para a interpretação dos dados.

Os modelos PLS-DA construídos a partir do espectro completo e com seleção de variáveis apresentaram desempenhos semelhantes, indicando que a redução do número de variáveis não comprometeu a capacidade discriminante dos modelos. Esse resultado demonstra que a seleção de regiões espectrais pode ser utilizada como estratégia para simplificação dos modelos, mantendo sua eficiência.

Para o café em grão, observou-se alto desempenho no conjunto de calibração, porém baixa capacidade de generalização no conjunto de validação externa, o que evidencia a influência das variações entre safras sobre a resposta espectral. Esse comportamento ressalta a importância de se considerar a variabilidade natural das amostras na construção de modelos mais robustos.

Com o objetivo de contornar essa limitação, foi desenvolvido um algoritmo iterativo para a obtenção de razões espectrais capazes de maximizar a separação entre os grupos e minimizar a variabilidade intragrupo. A equação final ajustada apresentou maior robustez frente às variações sazonais, possibilitando a discriminação das duas espécies mesmo para amostras provenientes de diferentes safras, além de permitir maior explicabilidade química.

De modo geral, os resultados obtidos indicam que a metodologia proposta apresenta potencial para aplicação em sistemas de autenticação e controle de qualidade de café, contribuindo para a prevenção de fraudes e para a valorização do produto.

REFERÊNCIAS

ABIC. **A Expansão do café no Brasil**. Disponível em: <<https://www.abic.com.br/tudo-de-cafe/a-expansao-do-cafe-no-brasil/>>. Acesso em: 22 jun. 2024.

ABIC. **Indicadores da Indústria de Café**. Disponível em: <<https://www.abic.com.br/estatisticas/indicadores-da-industria>>. Acesso em: 30 jun. 2025a.

ABIC. **Exportações Brasileiras de Café – Volume**. Disponível em: <<https://www.abic.com.br/estatisticas/exportacoes-brasileiras-de-cafe-volume>>. Acesso em: 19 jun. 2025b.

ABRANTES, Giovanna *et al.* Comparison between Variable-Selection Algorithms in PLS Regression with Near-Infrared Spectroscopy to Predict Selected Metals in Soil. **Molecules**, v. 28, n. 19, 1 out. 2023.

ADNAN, Adnan *et al.* Reliable discrimination of green coffee beans species: A comparison of UV-Vis-based determination of caffeine and chlorogenic acid with non-targeted near-infrared spectroscopy. **Foods**, v. 9, n. 6, 1 jun. 2020.

ALEIXANDRE-TUDO, J. L. *et al.* Chemometrics in food science and technology: A bibliometric study. **Chemometrics and Intelligent Laboratory Systems**, v. 222, 15 mar. 2022.

ALI, Khawla; AL-HAMEED, Abd. Spearman's correlation coefficient in statistical analysis. **Int. J. Nonlinear Anal. Appl**, v. 13, p. 2008–6822, 2022.

ATLABACHEW, Minaleshewa *et al.* Rapid and simultaneous determination of trigonelline, caffeine, and chlorogenic acid in green coffee bean extract. **Food Science and Nutrition**, v. 9, n. 9, p. 5028–5035, 1 set. 2021.

BEĆ, Krzysztof B.; GRABSKA, Justyna; HUCK, Christian W. **Near-infrared spectroscopy in bio-applications**. **Molecules** MDPI AG, , 1 jun. 2020.

BERVELIERI MADEIRA, Tiago *et al.* Desenvolvimento, validação e aplicação de um método rápido e sensível para determinação de trigonelina e cafeína por UHPLC-DAD em café torrado e moído. **Scientia Chromatographica**, v. 9, n. 3, p. 160–170, 2017.

BRASIL. **INSTRUÇÃO NORMATIVA Nº 8, DE 11 DE JUNHO DE 2003 - Regulamento Técnico de Identidade e de Qualidade para a Classificação do Café Beneficiado Grão Cru**. **MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO** BrasilDiário da União, Brasília, , 2003.

BRIANDET, R.; KEMSLEY, E. K.; WILSON, R. H. Approaches to Adulteration Detection in Instant Coffees using Infrared Spectroscopy and Chemometrics. **J Sci Food Agric**, v. 71, p. 359–366, 1996.

CARVALHO, André Cutrim *et al.* PANORAMA E IMPORTÂNCIA ECONÔMICA DO

CAFÉ NO MERCADO INTERNACIONAL DE COMMODITIES AGRÍCOLAS: UMA ANÁLISE ESPECTRAL. n. 9, p. 223–222, 2017.

CESTARI, Alexandre. Development of a fast and simple method to identify pure Arabica coffee and blended coffee by Infrared Spectroscopy. **Journal of Food Science and Technology**, v. 58, n. 9, p. 3645–3654, 2021.

CORREIA, Radigya M. *et al.* Portable near infrared spectroscopy applied to quality control of Brazilian coffee. **Talanta**, v. 176, p. 59–68, 1 jan. 2018.

COUTO, Cinthia de Carvalho *et al.* Near-infrared spectroscopy applied to the detection of multiple adulterants in roasted and ground arabica coffee. **Foods**, v. 11, n. 1, 1 jan. 2022.

COZZOLINO, Daniel. **The ability of near infrared (NIR) spectroscopy to predict functional properties in foods: Challenges and opportunities.** **Molecules**MDPI, , 1 nov. 2021.

DHARMAWAN, Agus; MASITHOH, Rudiati Evi. Authentication of Robusta Coffee Origins by Shortwave NIR Spectroscopy Coupled with Dimensionality Reduction and Neural Networks. **International Journal of Agriculture and Biosciences**, 2025.

EMBRAPA. **Rede Embrapa em Espectroscopia no Infravermelho Próximo - Net-NIRS.** Disponível em: <<https://www.embrapa.br/busca-de-projetos/-/projeto/38027/rede-embrapa-em-espectroscopia-no-infravermelho-proximo---net-nirs>>. Acesso em: 28 jun. 2024.

FERREIRA, Marcia Miguel Castro. **Quimiometria. Conceitos, Métodos e Aplicações.** Campinas: Editora Unicamp, 2015. v. 1

FORCHETTI, Débora A. P.; POPPI, Ronei J. Detection and Quantification of Adulterants in Roasted and Ground Coffee by NIR Hyperspectral Imaging and Multivariate Curve Resolution. **Food Analytical Methods**, v. 13, n. 1, p. 44–49, 1 jan. 2020.

FORNAZIER, Maurício José *et al.* Scale insects (Hemiptera: Coccoidea) associated with Arabica coffee and geographical distribution in the neotropical region. **Anais da Academia Brasileira de Ciências**, v. 89, n. 4, p. 3083–3092, 1 out. 2017.

GALVAN, Diego *et al.* APLICATIVO GAMMA-GUI: UMA INTERFACE GRÁFICA AMIGÁVEL PARA ANÁLISE EXPLORATÓRIA DE DADOS NO MATLAB. **Química Nova**, 2025.

GARCÍA MARTÍN, Juan Francisco. **Potential of Near-Infrared Spectroscopy for the Determination of Olive Oil Quality.** **Sensors**MDPI, , 1 abr. 2022.

JOLLIFE, Ian T.; CADIMA, Jorge. **Principal component analysis: A review and recent developments.** **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**Royal Society of London, , 13 abr. 2016.

KALSCHNE, Daneysa Lahis *et al.* Sensory characterization and acceptance of coffee brews of *C. arabica* and *C. canephora* blended with steamed defective coffee. **Food Research International**, v. 124, p. 234–238, 1 out. 2019.

KAMILOGLU, Senem. Authenticity and traceability in beverages. **Food Chemistry**, v. 277, p. 12–24, 30 mar. 2019.

KHARBACH, Mourad *et al.* **Current Application of Advancing Spectroscopy Techniques in Food Analysis: Data Handling with Chemometric Approaches.** **Foods** Multidisciplinary Digital Publishing Institute (MDPI), , 1 jul. 2023.

KY, C. L. *et al.* Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild *Coffea arabica* L. and *C. canephora* P. accessions. **Food chemistry**, p. 223–230, 11 abr. 2001.

MACHADO, Sandia *et al.* Assessing the differences of two vineyards soils' by NIR spectroscopy and chemometrics. **Heliyon**, v. 9, n. 12, 1 dez. 2023.

MARQUES, Catarina *et al.* **An Overview of Sensory Characterization Techniques: From Classical Descriptive Analysis to the Emergence of Novel Profiling Methods.** **Foods** MDPI, , 1 fev. 2022.

MEHMOOD, Tahir *et al.* **A review of variable selection methods in Partial Least Squares Regression.** **Chemometrics and Intelligent Laboratory Systems**, 15 ago. 2012.

MIHAILOVA, Alina *et al.* The use of multispectral imaging for the discrimination of Arabica and Robusta coffee beans. **Food Chemistry: X**, v. 14, 30 jun. 2022.

NATALE, Anna *et al.* Development of an analytical method for the determination of polyphenolic compounds in vegetable origin samples by liquid chromatography and pulsed amperometric detection at a glassy carbon electrode. **Journal of Chromatography A**, v. 1420, p. 66–73, 13 nov. 2015.

NGUYEN MINH, Quan *et al.* Authenticity green coffee bean species and geographical origin using near-infrared spectroscopy combined with chemometrics. **International Journal of Food Science and Technology**, v. 57, n. 7, p. 4507–4517, 1 jul. 2022.

PASQUINI, Celio. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. **J. Braz. Chem. Soc**, v. 14, n. 2, 2003.

PAULI, Elis Daiane *et al.* Detection of ground roasted coffee adulteration with roasted soybean and wheat. **Food Research International**, v. 61, p. 112–119, 2014.

RIBEIRO, J. S.; FERREIRA, M. M. C.; SALVA, T. J. G. Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. **Talanta**, v. 83, n. 5, p. 1352–1358, 15 fev. 2011.

SANDERS, Peter *et al.* The determination of sugars in dairy products: Development of

a new standard method for the international dairy federation and the internal organization for standardization. **Journal of AOAC International**, v. 100, n. 5, p. 1577–1581, 1 set. 2017.

SILVA, Edson Mario de Andrade *et al.* New genetic markers for 100% arabica coffee demonstrate high discriminatory potential for InDel-HRM-based coffee authentication. **Food Research International**, v. 173, 1 nov. 2023.

SIM, Joy *et al.* Non-linear machine learning coupled near infrared spectroscopy enhanced model performance and insights for coffee origin traceability. **Journal of Near Infrared Spectroscopy**, v. 32, n. 4–5, p. 127–139, 1 ago. 2024.

TOCI, Aline Theodoro *et al.* **Coffee Adulteration: More than Two Decades of Research. Critical Reviews in Analytical Chemistry** Taylor and Francis Ltd., , 3 mar. 2016.

TORNIAINEN, Jari *et al.* Open-source python module for automated preprocessing of near infrared spectroscopic data. **Analytica Chimica Acta**, v. 1108, p. 1–9, 29 abr. 2020.

VARELA, Paula; ARES, Gastón. **Sensory profiling, the blurred line between sensory and consumer science. A review of novel methods for product characterization. Food Research International**, out. 2012.

VOLSI, Bruno *et al.* The dynamics of coffee production in Brazil. **PLoS ONE**, v. 14, n. 7, 1 jul. 2019.

WANG, Hefei. Analysis of Brazilian Coffee Trade: Production, Prices, and Export Trends. 2025.

WANG, Xiuju; LIM, Loong Tak. Physicochemical Characteristics of Roasted Coffee. **Coffee in Health and Disease Prevention**, p. 247–254, 2015.

WATERS, Deborah M.; ARENDT, Elke K.; MORONI, Alice V. Overview on the mechanisms of coffee germination and fermentation and their significance for coffee and coffee beverage quality. **Critical Reviews in Food Science and Nutrition**, v. 57, n. 2, p. 259–274, 22 jan. 2017.

WEI, Feifei; TANOKURA, Masaru. Chemical Changes in the Components of Coffee Beans during Roasting. **Coffee in Health and Disease Prevention**, p. 83–91, 2015.

ZHANG, Dapeng *et al.* Accurate differentiation of green beans of arabica and robusta coffee using nanofluidic array of Single Nucleotide Polymorphism (SNP) markers. **Journal of AOAC International**, v. 103, n. 2, p. 315–324, 2020.

APÊNDICE A

Figura A.1 - Resultado PLS-DA utilizando o espectro pré-processado com SNV para o café cru moído e em grão

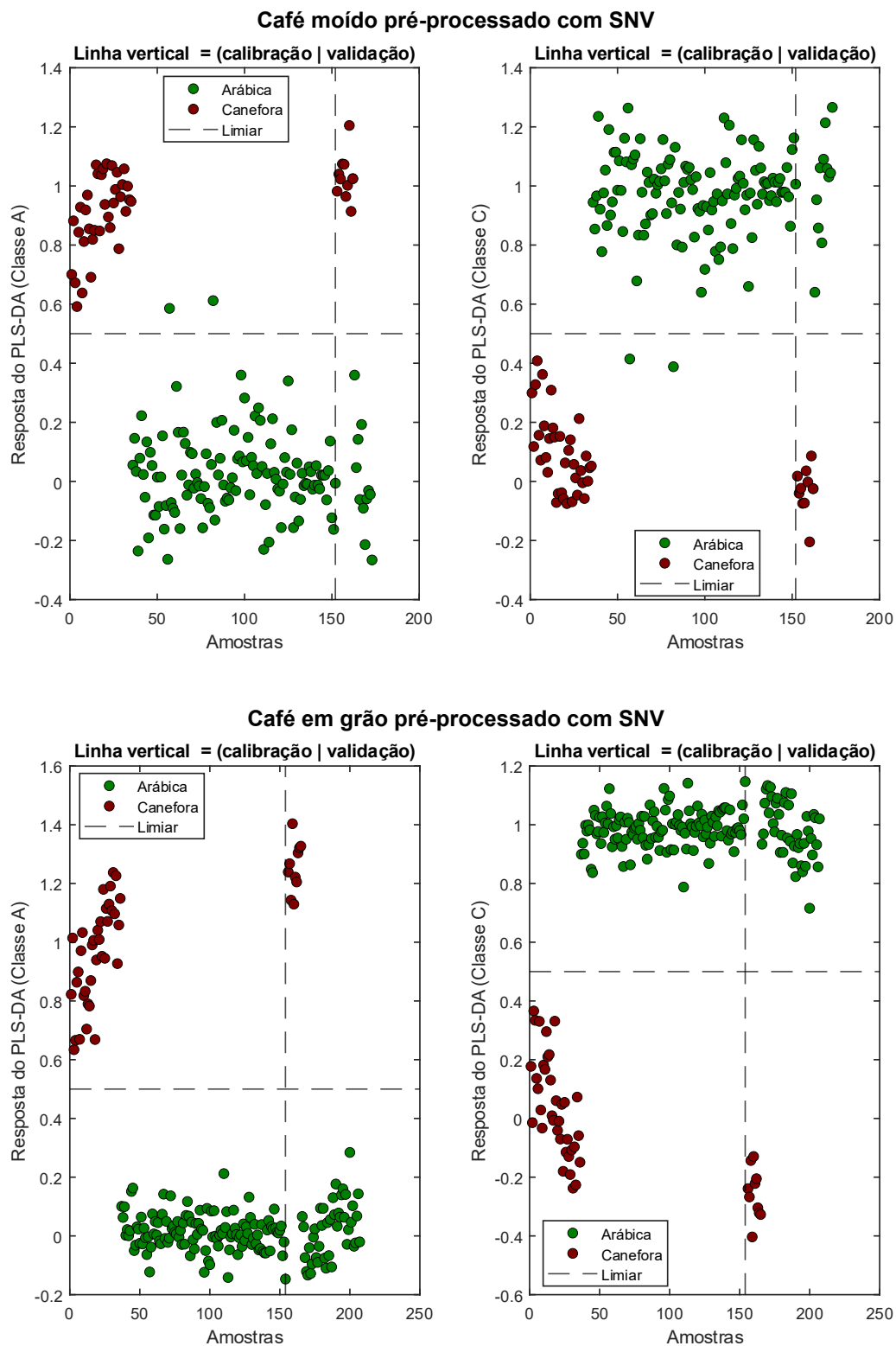
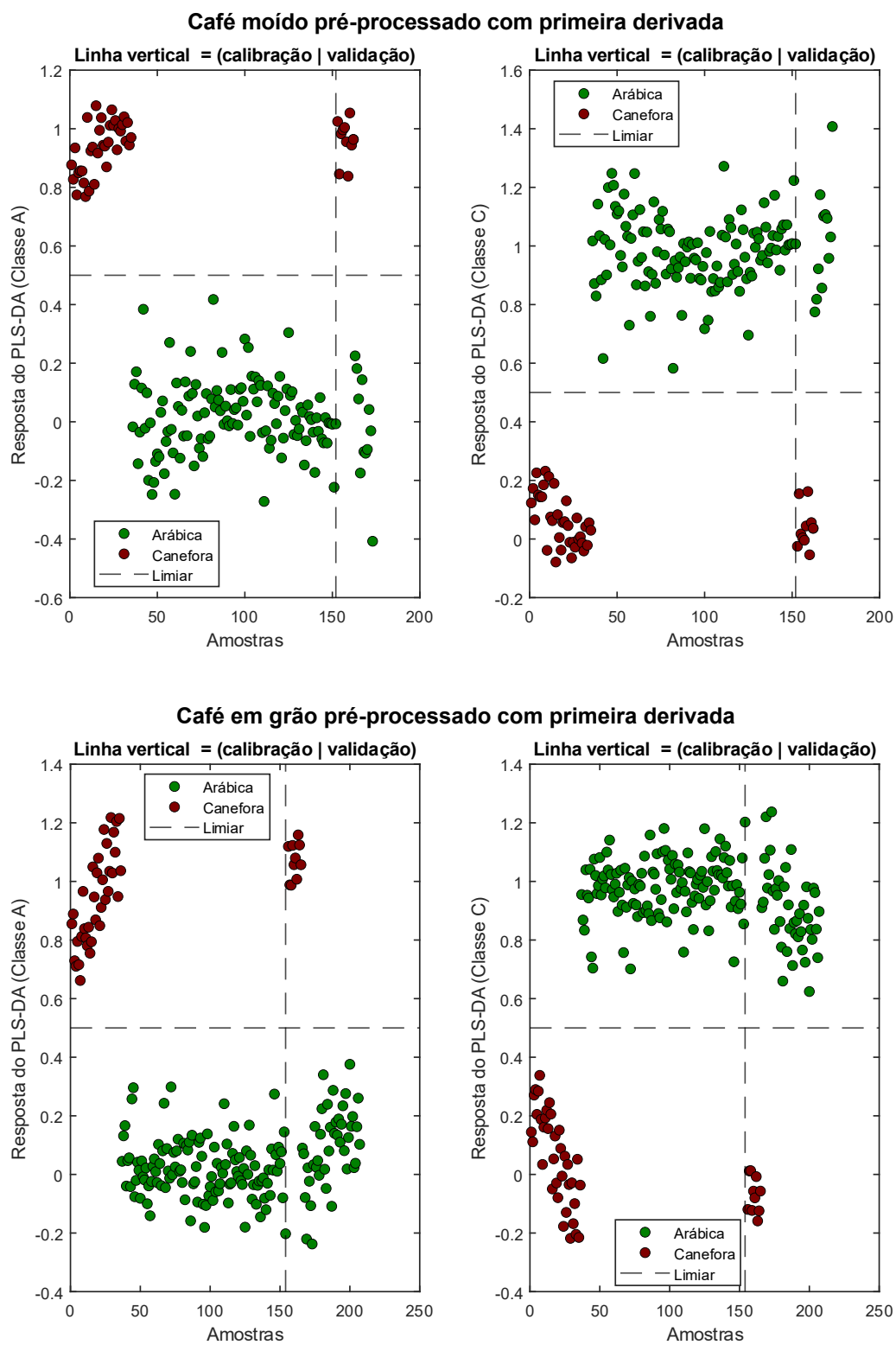


Figura A.2 - Resultado PLS-DA utilizando o espectro pré-processado com primeira derivada para o café cru moído e em grão



Fonte: o próprio autor

Tabela A.1 - Matriz de confusão do PLS-DA utilizando espectro pré-processado com SNV para o café cru moído e em grão para classe arábica

Amostra	Conjunto	VP	FN	FP	VN
Moído	Calibração	115	2	0	35
Moído	Validação	11	0	0	10
Grão	Calibração	118	0	0	36
Grão	Validação	43	0	0	10

Fonte: o próprio autor

Tabela A.2 - Matriz de confusão do PLS-DA utilizando espectro pré-processado com primeira derivada para o café cru moído e em grão para classe arábica

Amostra	Conjunto	VP	FN	FP	VN
Moído	Calibração	117	1	0	35
Moído	Validação	11	0	0	10
Grão	Calibração	118	0	0	36
Grão	Validação	43	0	0	10

Fonte: o próprio autor