



UNIVERSIDADE
ESTADUAL DE LONDRINA

VANESSA DE ALVARENGA

**APLICAÇÃO DA ESPECTROSCOPIA DE INFRAVERMELHO
PRÓXIMO ASSOCIADA A MÉTODOS QUIMIOMÉTRICOS PARA
PREDIÇÃO DE *BLENDS* DE *COFFEA ARABICA* E *COFFEA
CANEPHORA* EM GRÃOS CRUS DE CAFÉ**

Londrina

2023

VANESSA DE ALVARENGA

**APLICAÇÃO DA ESPECTROSCOPIA DE INFRAVERMELHO
PRÓXIMO ASSOCIADA A MÉTODOS QUIMIOMÉTRICOS PARA
PREDIÇÃO DE *BLENDS* DE *COFFEA ARABICA* E *COFFEA
CANEPHORA* EM GRÃOS CRUS DE CAFÉ**

Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Estadual de Londrina, como requisito parcial à obtenção do título de Mestre em Química.

Orientadora: Profa. Dra. Suzana Lucy Nixdorf

Londrina

2023

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

A473a Alvarenga, Vanessa de.

Aplicação da espectroscopia de infravermelho próximo associada a métodos quimiométricos para predição de blends de Coffea arabica e Coffea canephora em grãos crus de café / Vanessa de Alvarenga. - Londrina, 2023.
84 f. : il.

Orientador: Suzana Lucy Nixdorf.

Dissertação (Mestrado em Química) - Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em Química, 2023.
Inclui bibliografia.

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

CDU 54

VANESSA DE ALVARENGA

**APLICAÇÃO DA ESPECTROSCOPIA DE INFRAVERMELHO
PRÓXIMO ASSOCIADA A MÉTODOS QUIMIOMÉTRICOS PARA
PREDIÇÃO DE *BLENDS* DE *COFFEA ARABICA* E *COFFEA
CANEPHORA* EM GRÃOS CRUS DE CAFÉ**

Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Estadual de Londrina, como requisito parcial à obtenção do título de Mestre em Química.

BANCA EXAMINADORA

Profa. Dra. Suzana Lucy Nixdorf
Universidade Estadual de Londrina

Prof. Dr. Ednaldo José Ferreira
Empresa Brasileira de Pesquisa
Agropecuária

Profa. Dra. Diana Nara Ribeiro De Sousa
Universidade Estadual de Londrina

Londrina, 27 de julho de 2023.

**Aos meus queridos pais, que sempre me
incentivaram e apoiaram em todas as etapas
da minha vida.**

AGRADECIMENTOS

À Deus, por me dar forças e me ajudar a carregar as cruzes da minha vida.

À Mãe Rainha e Vencedora Três Vezes Admirável de Schoenstatt, mãe e educadora, por me guiar em todo o caminho, pelas graças recebidas e por todas as intercessões.

À Profa. Dra. Suzana Lucy Nixdorf, orientadora e amiga, pelos conhecimentos compartilhados, pelas contribuições no desenvolvimento desse trabalho e pelo companheirismo e incentivo ao longo desses anos.

Agradeço pelo fomento do presente trabalho realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, pela concessão da Bolsa, a qual foi fundamental para a continuidade da pesquisa.

Ao Instituto de Desenvolvimento Rural do Paraná – IAPAR – EMATER (IDR) e ao Ministério da Agricultura, Pecuária e Abastecimento (MAPA), pelo preparo e disponibilização das amostras de café analisadas nesse trabalho.

Ao Laboratório de Apoio à Pesquisa Agropecuária (LAPA) que compõe a Central Multiusuária de Laboratórios de Pesquisa da UEL (CMLP) pela utilização dos Equipamentos.

Ao meu grande amigo, Prof. Dr. Carlos Alberto Paulinetti da Camara, por todos os conselhos dados durante minha vida acadêmica, pelos conhecimentos compartilhados, pelas boas risadas no laboratório e por ser inspiração na minha carreira.

Ao meu amigo e parceiro de laboratório, Dr. Lycio Shinji Watanabe, por toda a ajuda no desenvolvimento desse trabalho, por compartilhar seu conhecimento, pela disposição em doar seu tempo para me ensinar coisas novas e, principalmente, pela paciência com minha pessoa.

À Universidade Estadual de Londrina, por, ao longo desses anos, prover os recursos para a minha aprendizagem, por ser minha segunda casa e ter me proporcionado momentos de alegria, superação e desenvolvimento pessoal.

Aos professores e funcionários que fazem parte do Centro de Ciências Exatas e aos integrantes do Grupo DIA (Desenvolvimento de Instrumentação, Automação e Metodologia Analítica), que contribuíram direta e indiretamente no desenvolvimento desse trabalho.

Ao meu namorado, Ronaldo Yoshikazu Fuji, que esteve me incentivando e apoiando ao longo de todo o mestrado.

Aos meus pais, Wanderley e Cleonice, pela paciência, incentivo e por me proverem recursos para que pudesse me dedicar exclusivamente aos estudos durante esses anos e aos meus irmãos, Vinicius e Viviane, por serem meus melhores amigos.

“Os homens deste mundo são como os grãos de café na máquina de moer: um antes, um depois, outro em seguida, todos acabam por ir para o mesmo destino.”

Giuseppe Gioachino Belli

ALVARENGA, Vanessa. **Aplicação da espectroscopia de infravermelho próximo associada a métodos quimiométricos para predição de blends de *Coffea arabica* e *Coffea canephora* em grãos crus de café.** 83 fls. Dissertação (Programa de Mestrado em Química) – Universidade Estadual de Londrina, 2023.

RESUMO

O café é um dos produtos mais apreciados pela população mundial, tendo sido consumidas em 2020 e 2021, 163 milhões de sacas no mundo. Esta *commodity* é relevante para o Brasil, levando em conta que o País, é seu maior produtor e exportador. Das 130 espécies da planta conhecidas, apenas duas possuem relevância comercial - *Coffea arabica* Lineu (café arábica) e *Coffea canephora* Pierre (café canéfora). Essas espécies diferem entre si, nas características físicas, químicas, sensoriais e em seu valor comercial. Embora sejam distintas, sua classificação manual é demorada. Tendo em vista que a matéria-prima define a qualidade do produto, justifica-se determinar sua identidade. Portanto, este trabalho objetivou desenvolver um método rápido utilizando a espectroscopia no infravermelho próximo (NIRS), para discriminar as espécies arábica e canéfora, e criar um modelo preditivo para determinar suas proporções nos *blends* a partir dos grãos crus. Os espectros NIRS foram obtidos diretamente de 50 g de café beneficiado grão cru, sem passar em peneira e nem triturar, dispostos no porta amostra, e lidos em 30 s na faixa de 750 a 2500 nm. Um banco de dados foi constituído pelas análises de 66 amostras da safra 2021/22 representativas de diferentes procedências nacionais, com espectros de 38 cafés arábica (MG, SP, PR, BA) e 28 cafés canéfora (RO e ES), submetidos aos pré-processamentos de correção multiplicativa de espalhamento (MSV). Empregando-se a análise de componentes principais (PCA), as componentes PC-1 e PC-2 foram capazes de juntas explicarem 96% da variância dos dados. A análise de agrupamento hierárquico (HCA) permitiu a separação em dois grandes grupos das 2 espécies, arábica e canéfora. Os *blends* de 50 g foram preparados a partir de 2 amostras, classificadas com número reduzido de defeitos, nas proporções de 5, 10, 20, 30, 40 e 50% (m/m) de café canéfora em arábica. Os gráficos de *scores* da PCA para os *blends* foram capazes de explicar 98% da variância dos dados. Para criar o modelo preditivo de distinção das misturas, utilizou-se a regressão por componentes principais (PCR) e a regressão por mínimos quadrados parciais (PLSR), com validação por meio do método *leave-one-out* (LOO). Embora os erros quadráticos médios (RMSE) foram elevados, os coeficientes de determinação foram adequados e os modelos preditivos PCR e PLSR foram capazes de estimar os valores das proporções de canéfora em arábica em relação ao referencial nas proporções acima de 40%. O método desenvolvido mostrou-se como alternativa verde e rápida de classificação das espécies de cafés, com potencial de predição de suas proporções em misturas, que tende a uma maior robustez com o acréscimo da variabilidade amostral.

Palavras-chave: grãos de café, NIRS, quimiometria, análise de componentes principais, método verde, modelos preditivos.

ALVARENGA, Vanessa. **Application of near infrared spectroscopy associated with chemometric methods to discriminate green coffee beans from *C. arabica* and *C. canephora* species.** 83 p. Chemistry Master's Degree Dissertation - State University of Londrina, 2023.

ABSTRACT

Coffee is one of the most appreciated products by the world's population, with 163 million bags consumed in the world in 2020 and 2021. This commodity is relevant to Brazil, considering that the country is the largest producer and exporter. Out of the 130 known species of the plant, only two have commercial relevance - *Coffea arabica* Lineu (arabica coffee) and *Coffea canephora* Pierre (canephora coffee). These species differ from each other in physical, chemical, sensory, and commercial characteristics. Although distinct, their manual classification is time-consuming and particularly challenging when blends are made for commercialization. Considering that the raw material defines the product quality, it is justified to determine its identity. Therefore, this study aimed to develop a rapid method using near-infrared spectroscopy (NIRS) to discriminate arabica from canephora species and create a predictive model to determine their proportions in blends based on raw beans. NIRS spectra were obtained directly from 50 g of raw processed coffee beans, without sieving or grinding, placed in the sample holder, and read within 30 s in the range of 750 to 2500 nm. A database was constituted by analyzing 66 samples from the 2021/22 harvest, representative of different national origins, with spectra of 38 arabica coffee (MG, SP, PR, BA) and 28 canephora coffee (RO and ES), subjected to the pre-processing of multiplicative scatter correction (MSC). Using principal component analysis (PCA), the PC-1 and PC-2 components were able to explain 96% of the data variance. Hierarchical cluster analysis (HCA) allowed the separation into two major groups of the two species, arabica and canephora. The 50 g blends were prepared from two samples, classified with a reduced number of defects, in proportions of 5, 10, 20, 30, 40, and 50% (m/m) of canephora coffee in arabica. The PCA score plots for the blends were able to explain 98% of the data variance. To create the predictive model for species distinction, Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR) were evaluated, validated using the leave-one-out (LOO) method. Although the mean squared errors (RMSE) were high, the coefficients of determination were adequate and the predictive models PCR and PLSR were able to estimate the values of the proportions of canephora in arabica in relation to the reference above 40%. The developed method proved to be a green and a fast alternative for classifying coffee species, with potential for predicting their proportions in mixtures, which tends to greater robustness with the increase in sample variability.

Key Words: coffee beans, NIRS, chemometrics, principal component analysis, green method, predictive models.

LISTA DE FIGURAS

Figura 1. Corte esquemático longitudinal do fruto do cafeeiro.....	20
Figura 2. Grão de café arábica (esquerda) e grão de café conilon (direita)	25
Figura 3. Gráfico esquemático da cinética da evolução do sabor do café durante a torrefação	29
Figura 4. Estruturas químicas dos polissacarídeos presentes nos grãos de café: (a) galactomananos, (b) arabinogalactanos e (c) celulose	31
Figura 5. Estrutura química do 5-o-cafeoilquínico	32
Figura 6. Estrutura química da (a) cafeína e (b) trigonelina	33
Figura 7. Principais bandas de absorção, com a frequência de grupos funcionais, na região do infravermelho próximo e sua localização no espectro magnético	37
Figura 8. Blends preparados com café arábica e canéfora nas proporções de 5, 10, 20, 30, 40 e 50% (m/m) em relação à canéfora	43
Figura 9. Cella do tipo “carinho” utilizada no espectrômetro de infravermelho para as amostras de grãos de café	45
Figura 10. Sobreposição da média da duplicata dos espectros NIRS para 66 amostras de grãos de cafés crus, sendo 38 de arábica e 28 de canéfora, com varredura de 750 a 2500 nm	47
Figura 11. Sobreposição da média da duplicata dos espectros NIRS para 66 amostras de grãos de cafés crus, sendo 38 de arábica e 28 de canéfora, com varredura de 750 a 2500 nm, com pré-processamento de MSC	48
Figura 12. Filtro Savitzky-Golay com derivadas e pré-processamento de MSC. (a) primeira derivada com janela de 9 pontos e polinômio de segunda ordem e (b) segunda derivada com janela de 7 pontos e polinômio de segunda ordem.....	49
Figura 13. Gráfico de scores, no plano bidimensional das componentes principais 1 e 2, referentes aos espectros na região do NIRS das 66 amostras de grãos de café cru das espécies café arábica (A) e café canéfora (C).....	51
Figura 14. Conjunto de espectros das 66 amostras de café e gráfico de loadings-X (linha azul) para a PC-1, com as variáveis do eixo X sendo os comprimentos de onda de 750 a 2500 nm	52
Figura 15. Gráfico de scores, no plano bidimensional da PC-1 e PC-2, referentes aos comprimentos de onda: (a) 750-900 nm; (b) 1680-1700 nm; (c) 1700-2000 nm e (d) 2000-2500 nm, do espectro NIRS	53
Figura 16. HCA referente às amostras de grãos de cafés crus das espécies café canéfora (C, em azul) e café arábica (A, em vermelho)	55
Figura 17. Sobreposição da média da duplicata dos espectros NIRS para 8 amostras de blends de grãos de cafés crus, com varredura de 750 a 2500 nm, para (a) espectro bruto e (b) espectro com pré-processamento de correção multiplicativa de espalhamento.....	56
Figura 18. Gráfico de scores, no plano bidimensional das componentes principais 1 e 2, referentes aos espectros na região do NIRS das 8 amostras de blends de grãos de cafés crus nas proporções 0, 5, 10, 20, 30, 40, 50 e 100% (m/m) de canéfora	57
Figura 19. Gráfico de valores Y referência versus Y previstos para o modelo PCR com PC-2, para 8 amostras de blends de grãos de café crus	60
Figura 20. Gráfico de valores Y referência versus Y previstos para o modelo PLS com Fator-2, para 8 amostras de blends de grãos de café crus	61

LISTA DE TABELAS

Tabela 1. Classificação do café beneficiado cru quanto à equivalência de defeitos intrínsecos	27
Tabela 2. Classificação do café beneficiado cru quanto à equivalência de impurezas extrínsecas	27
Tabela 3. Composição centesimal do grão de café cru e torrado das espécies <i>Coffea arabica</i> e <i>Coffea canephora</i>	30
Tabela 4. Regiões do espectro do infravermelho.....	36
Tabela 5. Composição dos blends de café arábica e canéfora utilizadas no estudo	44
Tabela 6. Classificação manual feita por classificador do MAPA seguindo a COB quanto ao número de defeitos em 300 g para os cafés arábica (SP099) e canéfora (RO006) utilizado para o preparo dos blends.....	44
Tabela 7. Parâmetros estatísticos dos conjuntos de calibração e validação por LOO, obtidos a partir do modelo PCR, para 8 amostras de blends de canéfora no arábica de 0, 5, 10, 20, 30, 40, 50 e 100% de grãos de cafés crus	58
Tabela 8. Parâmetros estatísticos dos conjuntos de calibração e validação por LOO, obtidos a partir do modelo PLSR, para 8 amostras de blends de canéfora no arábica de 0, 5, 10, 20, 30, 40, 50 e 100% de grãos de cafés crus	59
Tabela 9. Valores preditos pelos modelos PCR e PLSR e seus desvios para 8 amostras de blends de grãos de café crus, nas proporções de 0, 5, 10, 20, 30, 40, 50 e 100% de canéfora em arábica.....	61
Tabela 10. Amostras utilizadas no estudo, com dados de espécie, defeitos e tipo, segundo metodologia da COB	82

LISTA DE ABREVIATURAS E SIGLAS

¢	Centavos de dólar
ABIC	Associação Brasileira da Indústria do café
ACGs	Ácidos Clorogênicos
CECAFÉ	Conselho dos Exportadores de Café do Brasil
COB	Classificação Oficial Brasileira
CONAB	Companhia Nacional de Abastecimento
FIR	Infravermelho Distante (do inglês <i>Far Infrared</i>)
FTIR	Espectroscopia no Infravermelho por Transformada de Fourier (do inglês, <i>Fourier Transform Infrared Spectroscopy</i>)
FT-MIR	Espectroscopia no Infravermelho Médio por Transformada de Fourier (do inglês, <i>Fourier Transform Mid-Infrared Spectroscopy</i>)
HCA	Análise de Agrupamento Hierárquico (do inglês, <i>Hierarchical Cluster Analysis</i>)
HPLC	Cromatografia Líquida de Alta Eficiência (do inglês, <i>High Performance Liquid Chromatography</i>)
ISO	Organização Internacional de Normalização (do inglês <i>International Organization for Standardization</i>)
lb	Libra
LOO	<i>Leave-One-Out</i>
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
MIR	Espectroscopia de Infravermelho Médio (do inglês, <i>Mid-Infrared Spectroscopy</i>)
MLR	Regressão Linear Múltipla (do inglês, <i>Multiple Linear Regression</i>)
MSC	Correção de Espalhamento Multiplicativo (do inglês, <i>Multiplicative Scatter Correction</i>)
NIRS	Espectroscopia de Infravermelho Próximo (do inglês, <i>Near Infrared Spectroscopy</i>)
PCA	Análise de Componentes Principais (do inglês, <i>Principal Component Analysis</i>)

PCR	Regressão dos Componentes Principais (do inglês, <i>Principal Components Regression</i>)
PLSR	Regressão por Mínimos Quadrados Parciais (do inglês, <i>Partial Least Squares Regression</i>)
RMSE	Erro médio Quadrático (do inglês, <i>Root Mean Square Error</i>)
SCAA	Associação de Cafés Especiais da América (do inglês <i>Specialty Coffee Association of America</i>)
SVD	Decomposição de Valor Singular (do inglês <i>Singular Value Decomposition</i>)
USDA	Departamento de Agricultura dos Estados Unidos (do inglês, <i>United States Department of Agriculture</i>)
VBP	Valor Bruto da Produção

SUMÁRIO

1	INTRODUÇÃO	16
2	OBJETIVOS	19
2.1	OBJETIVO GERAL	19
2.2	OBJETIVOS ESPECÍFICOS	19
3	REVISÃO BIBLIOGRÁFICA	20
3.1	O CAFÉ	20
3.2	IMPORTÂNCIA ECONÔMICA DO CAFÉ NO BRASIL E NO MUNDO.....	22
3.3	AS ESPÉCIES COFFEA ARABICA E COFFEA CANEPHORA.....	24
3.4	CLASSIFICAÇÃO DO GRÃO DE CAFÉ CRU	26
3.5	COMPOSIÇÃO QUÍMICA DO CAFÉ	28
3.6	TÉCNICAS DE ANÁLISE DE CAFÉ	34
3.7	ESPECTROSCOPIA DE INFRAVERMELHO	35
3.7.1	ESPECTROSCOPIA DE INFRAVERMELHO PRÓXIMO	36
3.8	ANÁLISE QUIMIOMÉTRICA.....	38
3.8.1	ANÁLISE DE COMPONENTES PRINCIPAIS	39
3.8.2	ANÁLISE DE AGRUPAMENTO HIERÁRQUICO	39
3.8.3	REGRESSÃO DE COMPONENTES PRINCIPAIS	40
3.8.4	REGRESSÃO DE MÍNIMOS QUADRADOS PARCIAIS	41
4	METODOLOGIA	43
4.1	AMOSTRAS	43
4.2	PREPARAÇÃO DOS <i>BLENDS</i>	43
4.3	AQUISIÇÃO DOS ESPECTROS DE INFRAVERMELHO PRÓXIMO.....	44
4.4	ANÁLISES ESTATÍSTICAS	45
5	RESULTADOS E DISCUSSÃO	47
5.1	ESPECTROS DE INFRAVERMELHO PRÓXIMO.....	47
5.2	ANÁLISES QUALITATIVAS	50
5.2.1	ANÁLISE DE COMPONENTES PRINCIPAIS	50
5.2.2	ANÁLISE DE AGRUPAMENTO HIERÁRQUICO	54
5.2.3	AVALIÇÃO DO COMPORTAMENTO EM <i>BLENDS</i>	55
5.3	ANÁLISES QUANTITATIVAS	57
5.3.1	REGRESSÃO POR COMPONENTES PRINCIPAIS.....	58
5.3.2	REGRESSÃO POR MÍNIMOS QUADRADOS PARCIAIS.....	59

5.4	PREDIÇÃO DOS <i>BLENDS</i> A PARTIR DOS MODELOS	60
6	CONSIDERAÇÕES FINAIS	63
	REFERÊNCIAS	64
	ANEXO A	82

1 INTRODUÇÃO

O café é uma importante *commodity* agrícola, estando entre as mais comercializadas do mundo (FAO, 2021). É apreciado ao redor do globo, seja pelos benefícios conferidos à saúde ou pelo simples prazer de degustar sua bebida, que possui sabores e aromas marcantes (BOSSO; BARBALHO, 2021; SENINDE; CHAMBERS, 2020). Das 130 espécies de café conhecidas e suas diversas variedades (DAVIS; RAKOTONASOLO, 2021; DAVIS *et al.*, 2019; KORNMAN, 2019), duas possuem maior relevância comercial, as espécies *Coffea arabica* (café arábica) e *Coffea canephora* (café canéfora) (ICO, 2021).

De origem africana, as mudas de café chegaram ao Brasil no ano de 1727 e, pouco mais de um século depois, sua produção anual já girava em torno de 650 mil sacas (ICO, 2022; ALMEIDA *et al.* 2018; MARTINS, 2017). Atualmente, o café cru é o terceiro produto de maior renda nas exportações do setor agropecuário brasileiro, com um total de 115 países importadores, colocando o Brasil como maior exportador de café do mundo (ABIC, 2022; CNC, 2022). Além disso, o país é também o maior produtor de grãos de café, com plantações em 15 de seus estados, dos quais Minas Gerais, Espírito Santo, Bahia, São Paulo e Rondônia se destacam (CECAFÉ, 2022; ICO, 2022; BSCA, 2021; CONAB, 2021).

O consumo de café tem crescido nos últimos anos. Dados do Departamento de Agricultura dos Estados Unidos (USDA, do inglês *United States Department of Agriculture*), mostraram que foram consumidas 163 milhões de sacas de café no mundo entre os anos de 2020 e 2021, mesmo período em que ocorreu sua supervalorização no mercado internacional (CONAB, 2021; CONSÓRCIO PESQUISA CAFÉ, 2021; NYSE, 2021; USDA, 2021). Com consumidores cada vez mais exigentes, buscando novas experiências, a qualidade do café tem sido prioridade na hora de sua escolha (DOMINGUES; PEREIRA; BACIC, 2022; VENDRAME; FILHO, 2019). Atualmente existem selos que identificam a superioridade do produto, como o caso das certificações de pureza, qualidade e sustentabilidade criadas pela Associação Brasileira da Indústria de Café (ABIC) para oferecer aos consumidores transparência, qualidade e segurança (ABIC, 2022). A qualidade do café, identificada pelo seu sabor, está intrinsecamente relacionada à composição química dos grãos, que possui grande complexidade, e aos fatores que as influenciam, como espécie e variedade da planta, condições de pré e pós-colheita e grau de torrefação (LAMBOT

et al., 2016; MARCUCCI *et al.*, 2013; PIMENTA; VILELA, 2002).

O café das espécies *C. arabica* e *C. canephora* diferem-se em características físicas, químicas e sensoriais, sendo mais apreciados os cafés da espécie arábica por possuir notas mais suaves e maior dulçor (WANG; LIM, 2015; LIMA FILHO *et al.*, 2011). No mercado, são encontrados diversos tipos de café, desde os constituídos unicamente por arábica (cafés *gourmet* e especiais) até aqueles formados por misturas (*blends*) de arábica e canéfora (cafés superiores e tradicionais) em diferentes proporções (Resolução SAA – 28, 2007). Cada tipo de *blend* confere ao café mudanças em sua composição e em sua qualidade. Uma vez que o café arábica possui maior valor comercial, ele é por vezes alvo de adulterações, onde são adicionados café canéfora a sua composição para redução de custos sem as devidas informações de rotulagem, conseqüentemente lesando o consumidor. Sendo assim, a inspeção dos teores dos *blends* têm sido preocupação constante (FERREIRA *et al.*, 2021; SEZER *et al.*, 2018).

No intuito de normatizar o café produzido no Brasil há legislações que definem parâmetros de identidade e características mínimas de qualidade que grãos de cafés cru e torrado/moído devem obedecer. A portaria SDA nº 570, de 9 de maio de 2022 define as características de identidade e de qualidade para a classificação do café torrado/moído e a instrução normativa nº 8, de 11 de junho de 2003 estabelece o padrão oficial de classificação do café beneficiado grão cru (BRASIL, 2022; BRASIL, 2003). Tradicionalmente, para análise de qualidade utilizam-se: análise sensorial, feita pela prova de xícara por avaliadores sensoriais treinados pelo método da SCAA (*Speciality Coffee Association of America*); classificação dos defeitos, realizada com os grãos crus de forma manual, seguindo a Classificação Oficial Brasileira (COB); e análise de microscopia, que determina a pureza do café torrado e moído baseadas na observação microscópica (ABIC, 2022; SENAR, 2017; SCAA, 2015; LINGLE, 1993). Todas as análises citadas dependem de pessoas altamente qualificadas para obter resultados confiáveis, além de serem relativamente demoradas e inviáveis em análises de escala industrial. Atualmente, ainda não se tem uma metodologia oficial, rápida, para monitoramento da qualidade dos grãos de café cru.

Na literatura são encontrados técnicas e metodologias alternativas para verificação da qualidade de cafés e identificação de fraudes. Para a análise dos constituintes do café tem sido utilizadas técnicas analíticas como: cromatografia líquida de alta eficiência (CHEAH; FANG, 2020; DOMINGUES *et al.*, 2014;

MARCUCCI *et al.*, 2013; LUDWIG *et al.*, 2013; PAULI; VALDERI; NIXDORF, 2011; MUSSATTO *et al.*, 2011; ANDRADE *et al.*, 1998); cromatografia gasosa (LEE; SHIBAMOTO, 2002; CARRERA *et al.*, 1998; BICCHI *et al.*, 1997); cromatografias líquida e gasosa acoplada à espectroscopia de massas (OLIVEIRA *et al.*, 2009; ANDRZEJEWSKI *et al.*, 2004; MAEZTU *et al.*, 2001) e à ressonância magnética nuclear (CHARLTON *et al.*, 2002), espectroscopia de infravermelho próximo e médio (BELCHIOR; BOTELHO; FRANCA, 2022; CASTILLEJOS-MIJANGOS *et al.*, 2022; TOLESSA *et al.*, 2016; BARBIN *et al.*, 2014; SANTOS *et al.*, 2012; MORGANO *et al.*, 2007; ESTEBAN-DÍEZ; GONZÁLEZ-SÁIS; PIZARRO, 2004; DOWNEY *et al.*, 1997) e cromatografia líquida de alta eficiência com coluna de troca aniônica acoplada e detecção por amperometria pulsada (HPLC-HPAEC-PAD) (PAULI *et al.*, 2014; GARCIA *et al.*, 2009).

Dentre as técnicas analíticas citadas, a espectroscopia de infravermelho próximo (NIRS, do inglês *Near Infrared Spectroscopy*) vem ganhando espaço como uma alternativa rápida e de baixo custo, comparada com as demais, pois possui a vantagem de analisar pequenas quantidades de amostras com pouco ou nenhum preparo, além de ser uma técnica não destrutiva (CRAIG; FRANCA; OLIVEIRA, 2011; BARTON, 1987; OSBORNE; FEARN, 1986). A grande quantidade de informações obtidas por meio da espectroscopia de infravermelho faz com que a técnica requeira a utilização de métodos quimiométricos para auxiliar na interpretação dessas informações (BARBIN *et al.*, 2014; WOLD, 1995).

Estudos mostraram que a aplicação do NIRS em cafés torrado e moído para discriminação de espécies e detecção de adulteração tem sido eficaz (COUTO *et al.*, 2022; MUNYENDO; NJOROGÉ; HITZMANN, 2022; CORREIA *et al.*, 2018; BARBIN *et al.*, 2014; ESTEBAN-DÍEZ *et al.*, 2007; PIZARRO; ESTEBAN-DÍEZ; GONZÁLEZ-SÁIS, 2007). Todavia, são poucos os estudos que exploram essa técnica em grãos de cafés cru, principalmente quando se fala em *blends* dessa natureza (BAQUETA *et al.*, 2023; ADNAN *et al.*, 2020; GIRAUDDO *et al.*, 2019; TOLESSA *et al.*, 2016; DOWNEY; BOUSSION; BEAUCHÊNE, 1994). Considerando essa defasagem para o café cru, o presente trabalho teve como objetivo desenvolver um método rápido, que utilize a análise de espectroscopia no infravermelho próximo associado a quimiometria, para discriminar grãos de cafés crus das espécies arábica e canéfora (conilon), com o mínimo de preparo de amostras, e criar um modelo preditivo capaz de identificar a proporção dessas espécies diretamente em *blends* de café cru.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Desenvolver um método espectroscópico no infravermelho próximo por análise direta de café beneficiado grão cru, associado a quimiometria, para distinguir as espécies puras de café arábica e canéfora e criar um modelo preditivo para estimar a proporção da mistura nos *blends*.

2.2 OBJETIVOS ESPECÍFICOS

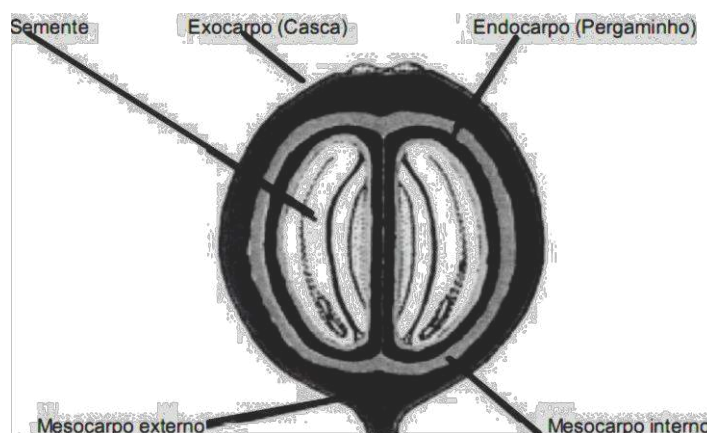
- Avaliar o potencial do método desenvolvido baseado nos espectros de infravermelho próximo, sem preparo prévio de amostra, juntamente com a aplicação da análise de componentes principais e análise de agrupamento hierárquico, em distinguir diretamente grãos crus de café arábica e canéfora.
- Criar um banco de dados no espectrômetro NIRS de bancada de grãos do perfil dos cafés crus *in natura* de diferentes regiões do Brasil das espécies arábica (MG, SP, PR e BA) e canéfora (ES e RO) para auxiliar no controle de qualidade do café.
- Construir modelos preditivos empregando Regressão por Componentes Principais (PCR) e Regressão por Mínimos Quadrados Parciais (PLSR), validando-os com método *leave-one-out* (LOO), visando estimar *blends* de grãos de café canéfora no arábica em diferentes proporções (5, 10, 20, 30, 40 e 50% (m/m)), verificando a porcentagem de acerto/erro em relação ao valor referencial.

3 REVISÃO BIBLIOGRÁFICA

3.1 O CAFÉ

O cafeeiro é uma árvore arbustiva nativa do continente africano pertencente à classe das dicotiledôneas e é por meio das sementes de seu fruto que se obtém o café (ICO, 2022; FERRÃO *et al.*, 2017). O fruto do cafeeiro (Figura 1) é popularmente conhecido como cereja por apresentar coloração vermelha, ou amarela, a depender da cultivar, em sua fase final de maturação. Seu fruto é do tipo drupa elipsoide, sendo composto por dois lócus, cada qual com uma semente, que por sua vez é recoberto pelo endocarpo, comumente chamado de pergaminho, e na parte mais externa pelo mesocarpo e pelo exocarpo, ou casca (HERRERA; LAMBOT, 2016; ALVES, 2007; AVALLONE *et al.*, 2000).

Figura 1. Corte esquemático longitudinal do fruto do cafeeiro.



Fonte: AVALLONE *et al.*, 2000.

Pertencente à família *Rubiaceae*, do gênero *Coffea* L., o cafeeiro possui uma ampla variedade de espécies. Até 2019, tinham-se conhecimento de 124 espécies de café, com sua maioria sendo encontrada em Madagascar, até que, em 2021, Davis e Rakotonasolo (2021) nomearam 6 novas espécies, subindo para o número de 130 conhecidas atualmente (DAVIS *et al.*, 2019; KORNMAN, 2019). As espécies com maior relevância econômica são duas: a *Coffea arabica* Lineu (café arábica), que equivale à 60% da produção mundial, e a *Coffea canephora* Pierre (café canéfora variedades conilon e robusta), correspondendo a 40% (ICO, 2021). Outras duas de menor cultivo, exploradas em certos países, são *Coffea liberica* (café libérica) e *Coffea*

dewevrei (café excelsa), as demais possuem importância no melhoramento genético da planta (FERRÃO *et al.*, 2017; DAVIS *et al.*, 2011; EMBRAPA, 2004).

Apesar de ter origem na Etiópia, foi na península Arábica que ocorreram as primeiras técnicas de plantio e preparo do café (RÁMON, 2014). Os árabes permaneceram com o monopólio até 1658, quando os holandeses conseguiram se apropriar de algumas mudas e iniciar o cultivo sistemático em Java (Indonésia), onde obtiveram sucesso (MARTINS, 2017). Nessa época, diversas casas de café já haviam sido abertas em países da Europa e popularizado a bebida, aumentando o interesse dos países em adquirir mudas para o plantio. Em pouco tempo, facilitado pela Companhia Holandesa das Índias Ocidentais, as mudas de café chegaram a várias regiões do globo, e em 1750 o cafeeiro já se encontrava nos cinco continentes (MARTINS, 2017; PENDERGRAST, 2010).

No Brasil, as primeiras mudas do cafeeiro vieram em 1727, trazidas da Guiana Francesa pelo sargento-mor Francisco de Melo Palheta. Elas foram semeadas no estado do Pará, entretanto, ficaram resumidas à pequenas áreas que não apresentaram desenvolvimento expressivo. Este fato resultou na migração da planta para outros estados que poderiam ter melhor adaptação, uma vez que as condições climáticas do Brasil se mostravam favoráveis ao cultivo (ALMEIDA *et al.* 2018; MARTINS, 2017; RUFINO, 2006). Pouco mais de um século depois, em 1830, já com plantações nas regiões do centro-oeste, sudeste e sul, o café se tornava o principal produto de exportações do país, resultado da alta procura do produto pelo mercado europeu e americano, e com uma produção anual em torno de 650 mil sacas, aproximadamente 26% de toda a produção cafeeira mundial (MARTINS, 2017; OCAMPO, 1984 *apud* RUFINO, 2006; TOPIK, 1987).

Dada tamanha importância para o Brasil, em 1850 se iniciou o chamado ciclo do café, que provocou grandes mudanças políticas e econômicas, enriquecendo os grandes fazendeiros produtores de café e promovendo a urbanização e o desenvolvimento do país, mas se beneficiando da estrutura escravista presente na época (ABIC, 2022). Em 1930, devido à Grande Depressão que abalou a economia mundial, o ciclo do café se deu por encerrado, com o setor industrial tomando a frente no desenvolvimento econômico nacional (ABIC, 2021).

Atualmente, o café cru é o terceiro produto de maior renda nas exportações do setor agropecuário brasileiro, atrás da soja e do milho (BRASIL, 2022), e, apesar de ter deixado de ser o produto principal de exportação do Brasil, o país continua sendo

o maior produtor e exportador do grão no mundo, título que detém há mais de 150 anos (ABIC, 2022; CNC, 2022; ALMEIDA *et al.*, 2018; ALBUQUERQUE, 1977). No país, as lavouras de café estão presentes em 15 de suas 27 unidades federativas, com destaque para os estados de Minas Gerais, Espírito Santo, Bahia, São Paulo e Rondônia, que concentram a maior parte da produção (CECAFÉ, 2022; ICO, 2022; BSCA, 2021).

3.2 IMPORTÂNCIA ECONÔMICA DO CAFÉ NO BRASIL E NO MUNDO

O café é uma importante *commodity* agrícola, de grande valorização comercial, que tem como principais produtores os países em desenvolvimento, contribuindo consideravelmente para suas receitas de exportações (ICO, 2022; NYSE, 2021; DOMINGUES *et al.*, 2014; PENDERGRAST, 2009).

Os principais produtores de café são Brasil, Vietnã, Colômbia e Indonésia, que correspondem em média a 35, 18, 8 e 7% de toda a produção mundial respectivamente, sendo predominante a produção de canéfora, variedade robusta, no Vietnã e Indonésia, e de arábica no Brasil e Colômbia (ICO, 2022).

Sobre a produção cafeeira no Brasil, o levantamento da safra de 2021 realizada pela Companhia Nacional de Abastecimento (CONAB) aponta que foram produzidas 47,7 milhões de sacas de café beneficiado, representando uma diminuição de 24,4% em relação à safra anterior, que bateu o recorde de produção. Desse total, 31,4 milhões sacas são de café arábica, que corresponde a um decréscimo de 35,5% na produção e os outros 16,3 são de café canéfora, que comparada ao ano anterior teve um incremento de 13,8%. Essa diminuição na produção cafeeira ocorreu em virtude de adversidades climáticas, como a seca e a geada, que provocaram queda no rendimento das lavouras e diminuição da área destinada à produção e pela bialidade negativa da safra (BRAINER; XIMENES, 2021; CONAB, 2021; MAPA, 2021).

No território brasileiro são em torno de 1,8 milhões de hectares em produção de café, sendo 81% dessa área destinada as lavouras de café arábica, que se concentram no estado de Minas Gerais. No Espírito Santo é onde se localiza a maior produção de café canéfora (CONAB, 2021).

Entrando na área de exportações, conforme o relatório mensal de junho de 2021 do Conselho dos Exportadores de Café do Brasil (CECAFÉ), a safra de 2020/21

bateu o recorde mundial de volume exportado pelo país, com um total de 45,6 milhões de sacas, que englobam café cru, torrado/moído e solúvel, representando um aumento de 13,3% em relação à safra de 2019/20, registrando valor bruto da produção em US\$ 5,8 bilhões (CECAFÉ, 2021; MAPA, 2021). São 115 países que importam o café brasileiro, tendo como principais parceiros comerciais os Estados Unidos da América, a Alemanha, a Bélgica, a Itália e o Japão.

O aumento considerável nas exportações é reflexo do crescente consumo da bebida pela população mundial. Segundo o USDA (2021), entre os anos de 2020 e 2021 foram consumidas mais de 163 milhões de sacas de café no mundo, das quais 26,9 milhões são dos Estados Unidos, que lidera o *ranking* de maior consumidor de café, seguido do Brasil, com 22,4 milhões de sacas (ICO, 2021). Somente no ano de 2021, o consumo *per capita* do brasileiro foi de 6,06 kg para o café cru, apontam os dados de Associação Brasileira da Indústria do café (ABIC, 2022), enquanto para o café torrado e moído foi de 4,84 kg, ambos com crescimento de 1,06% em relação ao ano anterior.

Dada tamanha procura pelo produto, acrescentada pelas adversidades climáticas dos últimos anos, o café sofreu alta dos preços no mercado internacional, contribuindo significativamente no faturamento das lavouras brasileiras (CONAB, 2021). Segundo dados do Consórcio Pesquisa Café (2021), no ano de 2021 o Valor Bruto da Produção (VBP) dos Cafés Brasil faturou R\$ 40,12 bilhões, sendo R\$ 30,7 bilhões apenas da espécie arábica, maior valor já registrado. Na Bolsa de Valores de Nova York, o preço médio do café chegou a ¢222,37/lb para o arábica e ¢102,33/lb para o robusta, representando um aumento de 97,6 e 63,9%, respectivamente, em relação ao mesmo período de 2020 (CONAB, 2021; NYSE, 2021).

Devido à alta valorização do café aliada à consumidores cada vez mais exigentes, a preocupação com a qualidade do produto tem sido foco de pesquisas ao longo dos anos, buscando meios de discriminar as espécies (OBEIDAT; HAMMOUDEH; ALOMARY, 2018; WEI *et al.*, 2012; BRIANDET; KEMSLEY; WILSON, 1996), verificar defeitos e adulterações (GIACALONE *et al.*, 2019; OLIVERI *et al.*, 2019; SEZER *et al.*, 2018; TOCI *et al.*, 2016; PAULI *et al.*, 2014), e analisar sua composição (RODRIGUES; BRAGAGNOLO, 2013; URYUPIN; PEREGUDOV, 2013), as quais estão diretamente relacionados com o sabor da bebida e preferências dos consumidores (ABIC, 2010; ARRUDA *et al.*, 2009; BORÉM, 2008).

3.3 AS ESPÉCIES *COFFEA ARABICA* E *COFFEA CANEPHORA*

As duas principais espécies de café no mercado, *C. arabica* e *C. canephora*, se diferenciam em características físicas, químicas e sensoriais (WANG; LIM, 2015; LIMA FILHO *et al.*, 2011; FONSECA *et al.*, 2007; ALVES, 2007; ILLY; VIANI, 2005; ICO, 1991).

De origem etíope, o café arábica foi a primeira espécie do gênero *Coffea L.* a ser descrita. Trata-se de uma planta autógama, que se reproduz por meio da autofecundação (SAKIYAMA; PEREIRA; ZAMBOLIM, 1999) e é a única do gênero *Coffea* considerada tetraploide, com 44 cromossomos ($2n = 44$). As demais, incluindo o conilon, são diploides, com 22 cromossomos ($2n = 22$) (ILLY; VIANI, 2005; MATIELLO, 1991).

O café é cultivado em regiões de alta altitude (900 a 2000 metros acima do nível do mar) e clima temperado (CORRÊA *et al.*, 2015). Comparado ao canéfora, o café arábica possui menor resistência a doenças e o seu rendimento após a torrefação é inferior. Em contrapartida, devido as diferenças na composição química, as bebidas produzidas com o café arábica possuem qualidade superior, são mais aromáticas e têm perfis adocicados, sendo utilizados para produzir cafés classificados como especiais e *gourmet* (AGNOLETTI *et al.*, 2019; ASSIS; OLIVEIRA; SENA, 2017; FARAH; SANTOS, 2015; LIMA FILHO *et al.*, 2013). Suas principais variedades são: típica, bourbon, caturra, mundo novo, catuaí e icatu (PATRÍCIO; BRAGHINI; FAZUOLI, 2010; FAZUOLI *et al.*, 2007; ANTHONY *et al.*, 2001).

O café canéfora tem origem no Congo e é popularmente conhecido como café robusta ou conilon, designação genérica a todas as variedades da *C. canephora* (MORAIS *et al.*, 2009). É uma planta alógama, que se reproduz por fecundação cruzada, cultivada em regiões de baixa altitude com clima quente e úmido, característico do clima equatorial (AGUIAR *et al.*, 2005).

Seus grãos são utilizados na industrialização do café solúvel, devido ao alto rendimento industrial advindo da concentração de sólidos solúveis dessa espécie, e na composição de misturas (*blends*) com arábica (MALAQUIAS; CELESTINO; XAVIER, 2018; MARTINS, 2007; FERNANDES *et al.*, 2003). A composição do *blend* de café influencia diretamente na qualidade da bebida e possui motivações estratégicas, uma vez que o valor de mercado do café arábica é superior ao canéfora, muitas vezes sendo alvo de adulterações no produto final (COUTO *et al.*, 2021;

COMBES; JOËT; LASHERMES, 2018; CORREIA *et al.* 2018; CORREIA *et al.*, 2016; PAULI; VALDERI; NIXDORF, 2011). Algumas das variedades mais comuns da espécie *C. canephora* são robusta, kouilou ou conilon, apoatã e guarini (FONSECA *et al.*, 2007).

Além das variedades de café mencionadas, existem também aquelas provenientes de melhoramento genético, chamados de híbridos. Os híbridos são cultivares criados a partir do cruzamento entre a própria espécie (intraespecífica) ou de espécies distintas (interespecífico), que é o caso do híbrido do timor, uma combinação de arábica e conilon que tem como característica resistência à ferrugem (MENDES *et al.*, 2007; KRUG; CARVALHO, 1952). Tais cruzamentos são estudados para aumentar a diversidade genética, criando variedades resistentes a doenças e estiagens, com maior rendimento e qualidade da bebida (ICO, 2022; LEROY *et al.*, 2006).

Fisicamente é possível notar algumas diferenças nos grãos de cada espécie. Como podemos ver na Figura 2 apresentada por Severini *et al.* (2017) o grão do café arábica é oval e de coloração verde claro, enquanto o do café conilon é arredondado e de coloração mais escura, tendendo a um tom castanho.

Figura 2. Grão de café arábica (esquerda) e grão de café conilon (direita)



Fonte: SEVERINI *et al.*, 2017.

Em pequenas quantidades, com grãos minimamente defeituosos, é possível fazer uma separação visual de acordo com a espécie, porém quando se trata de cafés torrados, moídos, *blends* ou até mesmo grãos crus com defeitos e em grande volume, as características físicas não são suficientes para a identificação, fazendo necessário o uso de técnicas para discriminação dessas espécies (ČÍŽKOVÁ *et al.*, 2007).

Na literatura são encontradas diferentes técnicas para essa finalidade, em sua grande maioria aplicadas a cafés torrado e moído (SAHACHAIRUNGRUENG *et al.*, 2022; CORREIA *et al.*, 2018; ESTEBAN-DÍEZ *et al.*, 2007; PIZARRO; ESTEBAN-

DÍEZ; GONZÁLEZ-SÁIS, 2007). Munyendo, Njoroge e Hitzmann (2022) apresentaram em seu trabalho as técnicas mais utilizadas para predição de espécies de café, mostrando que técnicas espectroscópicas (NIRS, MIR, FTIR) em conjunto com a quimiometria possuem grande potencial, mas ainda carecem de estudos com a variabilidade das condições amostrais. Estudos realizados por Downey, Boussion e Beauchêne (1994) por meio do NIRS, para cafés em grãos e moído, apresentaram bons resultados para a predição das espécies em cafés puros, com 96,20% de taxa de classificação, contudo, para os *blends*, essa taxa ficou entorno dos 83,02%.

3.4 CLASSIFICAÇÃO DO GRÃO DE CAFÉ CRU

A classificação do café ajuda a distinguir aqueles considerados de boa qualidade daqueles inferiores. Ela se baseia nas características físicas dos grãos, seguindo a Tabela Oficial Brasileira de Classificação e na análise sensorial da bebida, realizada por meio da prova de xícara de acordo com a SCAA e/ou pelo método da ABIC (ABIC, 2022; SCAA, 2015; BRASIL, 2003). O café beneficiado grão cru é classificado de acordo com: espécie (categoria); formato do grão e granulometria (subcategoria); sabor e aroma (grupo); tipo da bebida, sendo estritamente mole, mole, apenas mole, riado, rio e rio zona para bebidas de arábica; e excelente, boa, regular e ruim para bebidas de canéfora (subgrupo); coloração do grão (classe) e porcentual de defeitos e impurezas (tipo) (BRASIL, 2003).

A classificação quanto ao tipo avalia defeitos e impurezas que são perceptíveis na aparência (BARTHOLLO; GUIMARÃES, 1997). Os defeitos podem ser intrínsecos, aqueles que possuem imperfeições, ou seja, diferem da aparência de um grão sadio, e/ou extrínsecos, referente às impurezas encontradas nas amostras de café, em outras palavras, tudo aquilo que não for inerente ao grão (IBC, 1981). Para isso é possível encontrar na legislação a maneira correta de identificá-los (BRASIL, 2003).

A Tabela Oficial Brasileira de Classificação admite sete tipos de defeitos para o grão cru, que estão apresentados na Tabela 1. Para as impurezas (Tabela 2), são classificados quatro tipos, nesse caso são levados em consideração o tamanho das impurezas.

Tabela 1. Classificação do café beneficiado cru quanto à equivalência de defeitos intrínsecos

Defeitos	Quantidade	Equivalência
Grão Preto	1	1
Grão Ardido	2	1
Concha	3	1
Grão Verde	5	1
Grão Quebrado	5	1
Grão Brocado	2 a 5	1
Grão Mal Granado ou Choco	5	1

Fonte: BRASIL (2003)

Tabela 2. Classificação do café beneficiado cru quanto à equivalência de impurezas extrínsecas

Impurezas	Quantidade	Equivalência
Coco	1	1
Marinheiro	2	1
Pau, pedra, torrão grande	1	5
Pau, pedra, torrão regular	1	2
Pau, pedra, torrão pequeno	1	1
Casca grande	1	1
Casca pequena	2 a 3	1

Fonte: BRASIL (2003)

Por meio das Tabelas 1 e 2, é possível classificar os grãos por tipo e, caso a soma dos defeitos seja maior que 360, o café é considerado fora do tipo e pode ser comercializado desde que seja identificado na sua rotulagem essa informação. Com base nessa classificação é determinado a qualidade do produto e por sua vez, o seu valor comercial (BRASIL, 2022).

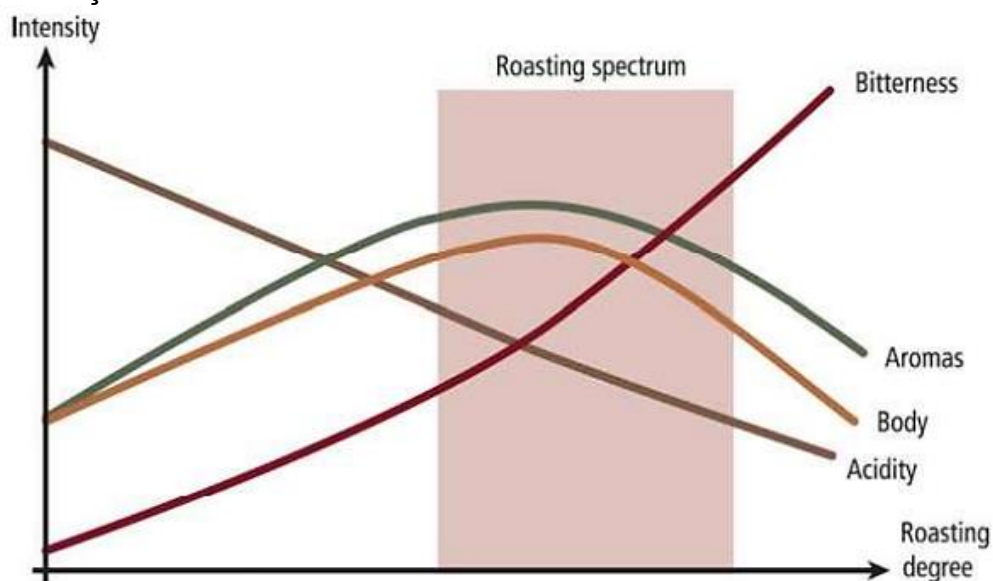
As classificações por meio da SCAA e da ABIC se dá por meio do grão torrado, avaliando a qualidade de sua bebida. A primeira categoriza o café em especial e não especial, já a segunda, possui quatro categorias, sendo elas: não recomendável, tradicional/extraforte, superior e gourmet (ABIC, 2022; SCAA, 2015).

3.5 COMPOSIÇÃO QUÍMICA DO CAFÉ

O café possui uma composição bastante complexa que depende de diversos fatores como a espécie e variedade da planta, o tipo de solo e seus nutrientes, sua localização (altitude), as condições edafoclimáticas a que a planta foi submetida no período de desenvolvimento e aos processos pós-colheita dos grãos. Esses últimos dependem da maturação, do manejo, das condições de armazenamento e, para os grãos torrados, do grau de torrefação (clara, média, escura) (LAMBOT *et al.*, 2016; PERRONE *et al.*, 2008; PIMENTA; VILELA, 2002).

O grão torrado possui diferenças em sua composição em relação ao grão cru devido às transformações químicas e físicas ocasionadas pelas altas temperaturas do processo da torra (SCHOLZ *et al.*, 2011; CLARKE; VITZTHUM, 2001). Nesse processo ocorre o rompimento da parede celular do grão, onde são liberados os compostos voláteis que dão origem ao aroma característico do café (ARYA; RAO, 2007; BUFFO; CARDELLI-FREIRE, 2004). O nível da torra influencia diretamente na bebida (MELO, 2004). Em torras claras é perceptível perfis doces, frutado, floral, de pão e nozes, produzindo cafés menos encorpados, finos e suaves, enquanto em torras escuras são característicos os perfis cacau, picante, fenólico e pungente (POISSON *et al.*, 2017). Em geral, a acidez diminui durante a torrefação enquanto o amargor é aumentado consideravelmente (Figura 3), principalmente em torras excessivas onde começam a aparecer notas tostadas e queimadas (SCHENKER; ROTHGEB, 2016). Já em torras média a acidez, amargor, aroma e corpo são equilibrados, nessa etapa o café já passou pelo processo de caramelização dos açúcares e começa a apresentar maiores níveis de amargor, mas ainda preservando uma certa doçura e notas aveludadas (CLIFFORD, 1999; SIVETZ; DESROSIER, 1979).

Figura 3. Gráfico esquemático da cinética da evolução do sabor do café durante a torrefação



Fonte: POISSON *et al.*, 2017.

Ainda no processo da torrefação, ocorrem reações intermediárias que transformam a sacarose e os aminoácidos em compostos nitrogenados de alta massa molar, chamadas de melanoidinas, o que justifica o decréscimo significativo desses componentes na composição do café torrado. As melanoidinas são formadas pela reação de Maillard, uma reação não enzimática responsável pela coloração marrom do café e uma das contribuintes da percepção do amargor e adstringência da bebida (SCHENKER; ROTHGEB, 2016; OOSTERVELD; VORAGEN; SCHOLS, 2003; CORREIA, 1990).

Segundo os estudos realizados por Wand e Lim (2015) e Bicho *et al.* (2011), a composição centesimal de alguns componentes do café variam de acordo com sua espécie (Tabela 3).

Tabela 3. Composição centesimal do grão de café cru e torrado das espécies *Coffea arabica* e *Coffea canephora*

Componentes	<i>Coffea arabica</i> ^a		<i>Coffea canephora</i> ^a	
	Cru	Torrado	Cru	Torrado
Polissacarídeos	34 – 45	31 – 38	47 – 55	37 – 42
Sacarose	6,0 – 9,0	0 – 4,2	0,9 – 4,8	0 – 1,6
Açúcares redutores	0,1 – 0,8	0,3	0,31 – 0,4	0,3
Lipídios	15 – 18	17	8 – 12	11
Proteínas	8,5 – 12	7,5 – 10	8,5 – 15	7,5 – 10
Cafeína	0,8 – 1,4	1,1 – 1,3	1,7 – 2,4	2,4 – 2,5
Trigonelina	0,6 – 2,0	0,2 – 1,2	0,3 – 0,9	0,3 – 0,7
Ácidos clorogênicos	6,7 – 9,2	1,9 – 2,5	7,1 – 12,1	3,3 – 3,8
Ácidos alifáticos ^b	2 – 2,9	1,6	1,3 – 2,2	1,6
Lignina	3,0	3,0	3,0	3,0
Pectina	2,0	2,0	2,0	2,0
Minerais	3 – 5,4	4,5	3 – 5,4	4,7
Aminoácidos	0,2 – 0,8	0	0,8 – 1,0	0
Água	8 – 12	0 – 5	8 – 12	0 – 5
Melanoidinas	0	23 – 25	0	23 – 25

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

^bÁcidos alifáticos incluem os ácidos cítrico, málico e quínico

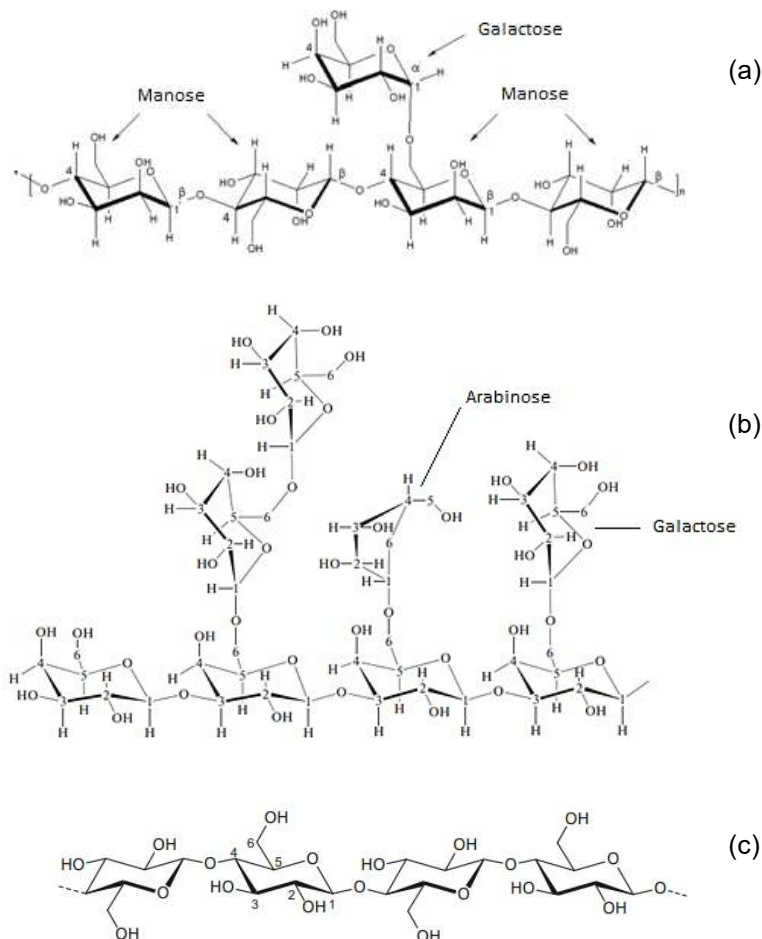
Fonte: Adaptado de Wang e Lim (2015), Bicho *et al.* (2011) e Clarke e Vitzthum (2001)

Comparando o grão cru nas duas espécies, observa-se que os grãos de arábica possuem teores mais elevados de sacarose, trigonelina e lipídios, enquanto os grãos crus de canéfora contêm mais polissacarídeos, cafeína, aminoácidos e ácidos clorogênicos (WANG; LIM, 2015; OESTREICH-JANZEN, 2013; SCHOLZ, 2008; FRANK; ZEHENTBAUER; HOFMANN, 2006). Dentre todos os componentes, quatro deles se destacam sendo comumente utilizados para diferenciação de cafés (CORREIA *et al.*, 2016), são eles: carboidratos (FISCHER *et al.*, 2001), ácidos clorogênicos (ACGs) (JESZKA-SKOWRON *et al.*, 2016), cafeína e trigonelina (CARACOSTEA; SIRBU; BUSURICU, 2021; DIAS; BENASSI, 2015; MARCUCCI *et al.*, 2013).

O café é composto em sua maior parte por carboidratos, representando mais de 50% de seu peso em base seca (WANG; LIM, 2015). Os carboidratos compõem majoritariamente a parede celular dos grãos, que é constituída por três principais

polissacarídeos: galactomananas, arabinogalactanas e celulose (Figura 4) (REDGWELL; FISCHER, 2006; FISCHER *et al.*, 2001; BUCKERIDGE *et al.*, 2000). As galactomananas e arabinogalactanas são polímeros de cadeia linear formadas por oligossacarídeos unidos por ligações glicosídeas, sendo galactose e manose para a primeira e arabinose e galactose para a segunda (OOSTERVELD *et al.*, 2003). Já a celulose é constituída por um único monômero, a glicose (LEHNINGER; NELSON; COX, 2014).

Figura 4. Estruturas químicas dos polissacarídeos presentes nos grãos de café: (a) galactomananos, (b) arabinogalactanos e (c) celulose



Fonte: Adaptado de: SECOUARD, GRISEL, MALHIAC (2007); b: MALYAR *et al.* (2018); c: HEINZE (2015)

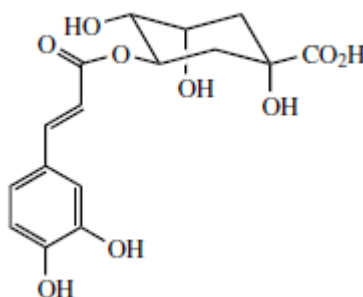
Os polissacarídeos contribuem para o corpo da bebida, formando complexos com proteínas e fenóis e interferindo diretamente em sua viscosidade. São conhecidos por estabilizar a espuma dos cafés expressos e reter os compostos aromáticos

(MOREIRA *et al.*, 2015; SALVA; LIMA, 2007; NUNES *et al.*, 2006; NUNES; COIMBRA, 1998). Já a quantidade de sacarose favorece a doçura da bebida, um dos motivos pelo qual o café arábica é mais apreciado, sendo mais suave e doce ao paladar, com notas florais e frutadas. No café beneficiado grão cru, a sacarose corresponde de 5 a 12% de seu peso, sendo precursor do aroma característico do café no processo da torra, onde vários compostos são gerados, como os ácidos carboxílicos (PERRONE *et al.*, 2008; OOSTERVELD; VORAGENS; SCHOLS, 2003).

Monossacarídeos como arabinose, galactose, manose, frutose, xilose e glicose são utilizados para identificação de fraudes em cafés solúveis, uma vez que são parâmetros que geralmente divergem quando adicionados materiais distintos ao produto (PORTILLO; ARÉVALO, 2022; CONSTANTINO *et al.*, 2020; PAULI; CRISTIANO; NIXDORF, 2011). Em cafés crus, a porcentagem de alguns dos monossacarídeos encontrados é de 0,03% de glicose, 0,04% de frutose e 5,0 a 12,0% de sacarose do peso seco (ROGERS *et al.*, 1999). A técnica para análise desses carboidratos em cafés solúveis é normatizada pela ISO 11292, que utiliza a cromatografia de troca aniônica acoplada com detecção eletroquímica por amperometria pulsada (HPAE-PAD, do inglês *High Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection*) para quantificá-los (ISO/11292, 1995).

Os ácidos clorogênicos são ésteres formados a partir da esterificação do ácido quínico com os ácidos caféico, cinâmico, ferúlico e p-cumárico. Entre os ácidos clorogênicos que compõem o café, o 5-o-cafeoilquínico (Figura 5) é o mais abundante e é a forma esterificada do ácido caféico com o ácido clorogênico (IWAI *et al.*, 2004; NATELLA *et al.*, 2002).

Figura 5. Estrutura química do 5-o-cafeoilquínico

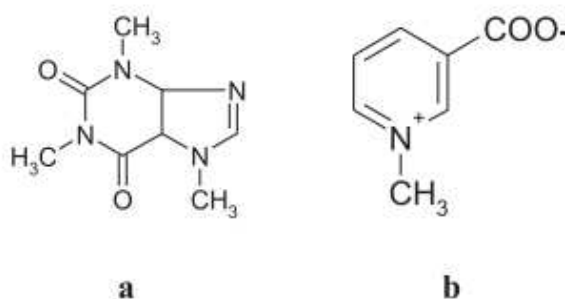


Fonte: DEWICK (2001)

Esses ácidos são degradados no processo da torra, reduzindo sua concentração e produzindo compostos fenólicos, proporcionando na bebida aromas de especiarias, cravo e sensação de adstringência (ABRAHÃO *et al.*, 2008; PERRONE *et al.*, 2008). No café arábica o teor desses ácidos é superior ao do café canéfora, impactando no seu valor comercial, uma vez que esses ácidos influenciam na formação do aroma durante o processamento do café e impactam na qualidade final do produto (MAESO *et al.*, 2006; SALDAÑA *et al.*, 1997).

A cafeína e trigonelina são dois importantes compostos nitrogenados presentes no café, que fazem parte dos alcaloides (MORAES *et al.*, 2008). A cafeína (Figura 6a), pertencente ao grupo das xantinas (MADEIRA *et al.*, 2017; MONTEIRO; TRUGO, 2005; ARNAUD, 1999), é o componente mais conhecido do café e possui estabilidade no processo da torrefação (NOGUEIRA; TRUGO, 2003). Ela é inodora e possui gosto amargo, além de trazer sensação de adstringência, características marcantes do café conilon que possui o dobro de cafeína em relação ao arábica (MONTEIRO; TRUGO, 2005). Ela está associada a ações biológicas no organismo humano, principalmente à estimulação do sistema nervoso central (CAZARIM *et al.*, 2014; IBRAHIM *et al.*, 2006). A trigonelina (Figura 6b), está presente em maiores concentrações em cafés crus pois é degradada no processo da torrefação, formando ácido nicotínico (niacina ou vitamina B3) e derivados da piridina. Esses compostos vão contribuir para o gosto amargo e pela intensidade de aroma (MAZZAFERA, 1991; MONTEIRO e TRUGO, 2005).

Figura 6. Estrutura química da (a) cafeína e (b) trigonelina



Fonte: MONTEIRO, TRUGO (2005)

Além das percepções sensoriais distintas entre as espécies devido às diferenças de concentração desses compostos, os quatro componentes explanados podem nos dar muitas informações sobre a espécie quando utilizado o NIRS, uma vez

que esses contêm grupos funcionais que geram bandas características no espectro de infravermelho (MUNYENDO; NJOROGE; HITZMANN, 2022). Variações de composição podem afetar sutilmente a forma espectral das amostras (KEMSLEV; RUAULT; WILSON, 1995).

3.6 TÉCNICAS DE ANÁLISE DE CAFÉ

Tradicionalmente, para análise de qualidade utilizam-se: análise sensorial; classificação dos defeitos seguindo a COB; e análise de microscopia (ABIC, 2022; SENAR, 2017; SCAA, 2015; LINGLE, 1993). A análises sensorial é a mais utilizada dentre elas, onde se tem avaliadores treinados que, por meio de uma metodologia conhecida como prova de xícara, avaliam os cafés e os classificam (ALVARADO; LINNEMANN, 2010; FEIRA-MORALES, 2002). Essa metodologia foi criada pela SCAA, mas como demanda uma alta qualificação do profissional sensorial, é relativamente demorada e possui limitações de quantidade de amostras, técnicas instrumentais de análise de café foram desenvolvidas, justamente para complementar essa técnica e levar o controle da qualidade à escala industrial.

Cheah e Fang (2020) utilizaram a cromatografia de alta eficiência, com análise quimiométrica, para identificar adulteração em café torrado, alcançando limite de detecção em torno de 5% na separação do café puro e adulterado. Já Pauli, Valderi e Nixdorf (2011) utilizaram da mesma técnica, com detector UV-Vis com pós-coluna, para determinação de carboidratos em cafés adulterados, onde foi observado variação dos teores de xilose e glicose em café com adulteração por derivados de cascas e milho, resultados positivos para aplicação em rotina para triagem de amostras.

Maeztu *et al.* estudaram aromas de cafés expressos de diferentes variedades botânicas por GC-MS e análise sensorial, obtendo sucesso na classificação das amostras segundo sua espécie. Propriedades físico-químicas, aliadas à quimiometria, também foram utilizadas para discriminação de espécies. Agnoletti e colaboradores (2019) desenvolveram um modelo capaz de distinguir as espécies arábica e conilon por meio do pH, acidez, compostos fenólicos totais e cafeína.

As técnicas espectroscópicas também são muito utilizadas em análises de qualidade do café, ganhando cada vez mais espaço pelas diversas vantagens com relação a outras, como uma alternativa mais rápida e barata. Castillejos-Mijangos *et al.* (2022) em seu trabalho apresentaram uma revisão dos estudos com aplicação da

técnica FT-MIR em cafés para diferentes finalidades, como detecção e quantificação de adulterantes, classificação geográfica e de genótipo e classificação por grau de torra. Marquetti *et al.* (2016) avaliaram o potencial do NIRS em distinguir origens geográficas de cafés arábica cru e Esteban-Díez e colaboradores (2006) utilizaram a mesma técnica para discriminar cafés torrados puro arábica e robusta e seus *blends*.

3.7 ESPECTROSCOPIA DE INFRAVERMELHO

A espectroscopia de infravermelho é uma técnica analítica instrumental que realiza medições das interações da radiação eletromagnética com a matéria, registrando a absorção, emissão e reflexão da energia, provendo informações sobre a constituição, estrutura molecular e interações entre as moléculas de determinado material (ZAREEF *et al.*, 2020; ALCÁCER, 2007).

A absorção ocorre quando a radiação infravermelha aplicada sobre o objeto se iguala a vibração natural das moléculas que o compõe. Essa energia absorvida faz com que os átomos das moléculas vibrem e registrem características em bandas no espectro, as quais estão diretamente relacionadas à estrutura dessas moléculas, como uma impressão digital (no inglês, *fingerprint*) (HOF, 2003; SILVERSTEIN; WEBSTER, 2000). Apenas moléculas que possuem momento de dipolo são capazes de absorver a radiação infravermelha, uma vez que para isso deve haver uma mudança em seu momento de dipolo para gerar um campo elétrico que interaja com o campo da radiação incidente, sendo assim, moléculas homonucleares não absorvem nessa região (PASQUINI, 2018; PASQUINI, 2003).

Existem dois tipos de vibrações moleculares, as de estiramento de ligação e as de deformação angular. Vibrações de estiramento são causadas quando ocorre a variação da distância entre átomos ligados, podendo ser simétrico ou assimétrico. Já a vibração de deformação angular ocorre quando essa variação se dá no ângulo de ligações adjacentes de um átomo da molécula (HIGSON, 2009; BARBOSA, 2008).

O espectro de infravermelho é dividido em três regiões de acordo com o intervalo de comprimento de onda (Tabela 4), sendo elas: infravermelho próximo (NIRS), infravermelho médio (MIR) e infravermelho distante (FIR, do inglês *Far Infrared*) (SKOOG; HOLLER; NIEMAN, 2009; BARBOSA, 2008; PAVIA; LAMPMAN; KRIS, 2001).

Tabela 4. Regiões do espectro do infravermelho

Região	Comprimento de onda (nm)	Número de onda (cm^{-1})	Frequência (Hz)
Próximo	750 a 2500	12800 a 4000	$3,8 \times 10^{14}$ a $1,2 \times 10^{14}$
Médio	2500 a 5000	4000 a 200	$1,2 \times 10^{14}$ a $6,0 \times 10^{12}$
Distante	5000 a 100000	200 a 10	$6,0 \times 10^{12}$ a $3,0 \times 10^{11}$

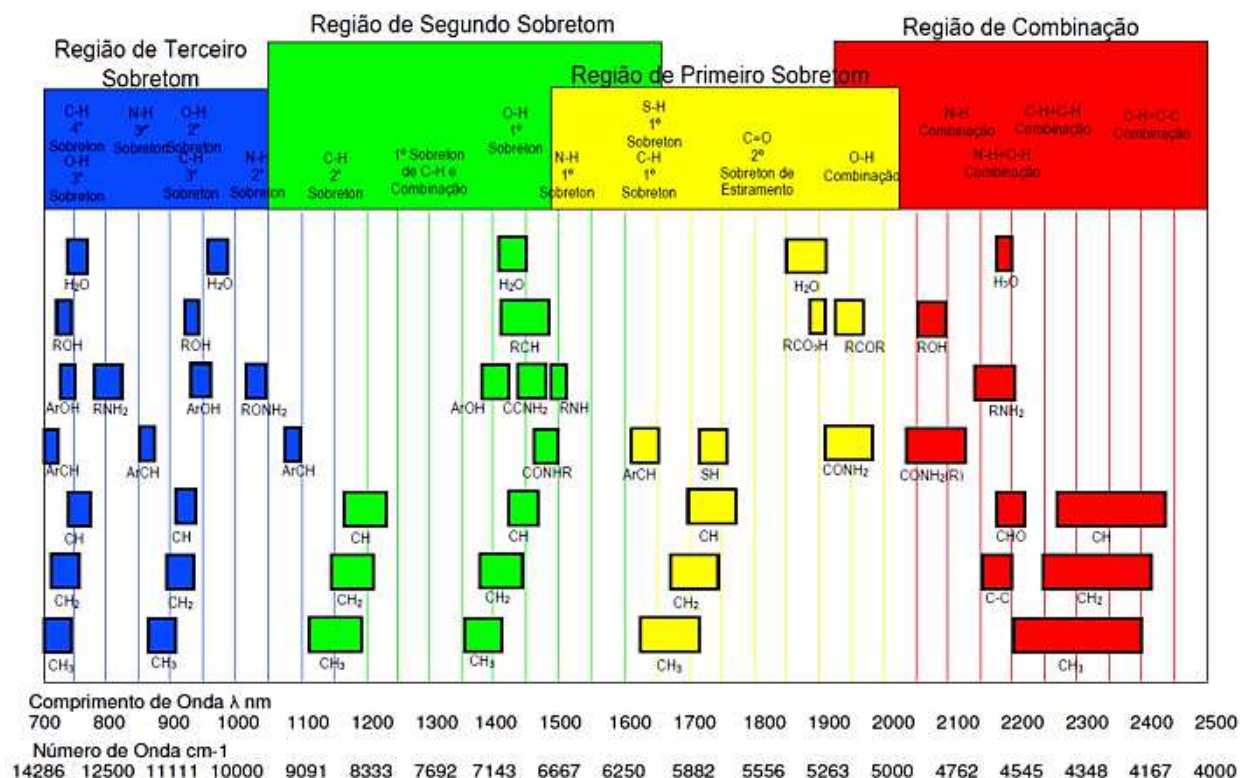
Fonte: SKOOG; HOLLER; NIEMAN (2009)

Os limites dos intervalos do espectro podem ser divergentes dependendo da literatura utilizada (PASQUINI, 2003). Para esse trabalho, foi dado enfoque para a região do infravermelho próximo.

3.7.1 ESPECTROSCOPIA DE INFRAVERMELHO PRÓXIMO

NIRS é uma técnica analítica baseada na espectroscopia vibracional que abrange a região de 750 a 2500 nm na faixa espectral (SKOOG; HOLLER; NIEMAN, 2009). As interações percebidas pela radiação nessa faixa são sobretons (*overtones*), combinações e ressonâncias dos modos vibracionais, principalmente de grupos funcionais que contenham átomos relativamente pesados como o carbono, nitrogênio, oxigênio e enxofre, que, por conterem dipolos intensos, apresentam absorção da radiação (PASQUINI, 2003). Moléculas orgânicas têm padrões de absorção específicos na região do infravermelho que podem relatar a sua composição. Na Figura 7 encontram-se as regiões de absorção de grupos funcionais para serem consultadas na realização das atribuições das bandas obtidas no espectro.

Figura 7. Principais bandas de absorção, com a frequência de grupos funcionais, na região do infravermelho próximo e sua localização no espectro magnético



Fonte: FERREIRA (2015) *apud* XIAOBO *et al.* (2010)

Para o café torrado e moído alguns autores reportaram as seguintes bandas de absorção no espectro de infravermelho:

- Tugnolo *et al.* (2021) definiram em 1450 nm o primeiro sobretom de estiramento –OH para água, em 1100 a 1250 nm o segundo sobretom de estiramento C-H para carboidratos, ácido quínico e lipídios, em 1300 a 1350 nm bandas de cafeína e em 1550 nm de ACGs, carboidratos e aminoácidos;
- Na região de 1850 a 1950 nm, Okubo e Kurata (2019) descreveram as bandas como primeiro sobretom de vibrações C=O de compostos como cafeína, lipídios, ACGs, proteínas, água e carboidratos;
- Marquetti e colaboradores (2016) atribuíram para lipídios as regiões de 1410, 1742, 1904 e 2318 nm, para a cafeína em 1410, 1728, 1904, 2306 e 2348 nm e para ACGs em 1436, 1880, 2312, 2324 e 2350 nm;
- Ribeiro, Ferreira e Salva (2011) retrataram as vibrações de combinação de bandas C-H + C-H e C-H + C-C e C-H do primeiro sobretom para lipídios nas regiões 1706-1704, 2436-2475 e 2480-2488 nm;
- Scholz *et al.* (2014) atribuíram em 1208 nm a região de estiramento C-H de

segundo sobretudo para sacarose, lipídios e aminoácidos e em 2300-2400 nm as combinações das bandas C-H + C-H₂ de moléculas de lipídios.

Os espectros do NIRS apresentam grande sobreposição de bandas e correlações entre as variáveis espectrais, o que dificulta a interpretação visual ou univariada dos dados (BRERETON, 2003b; PASQUINI, 2018). O desenvolvimento dos métodos quimiométricos se torna necessário para auxiliar na interpretação dessas informações por meio de construção de modelos capazes de descrever o comportamento estatístico observado (BARBIN *et al.*, 2014; WOLD, 1995).

3.8 ANÁLISE QUIMIOMÉTRICA

A quimiometria surgiu na década de 70 como uma área de estudo que envolve conhecimentos químicos, estatísticos e computacionais, tendo como pioneiros o Professor Doutor Bruce Kowalski, da Universidade de Washington (EUA) e o Professor Doutor Svante Wold, da Universidade de Umea (Suécia) (BRERETON, 2018). Essa área da química possui o objetivo de processar os dados experimentais por meio de ferramentas computacionais e extrair as informações mais relevantes, a fim de diminuir a extensão dos resultados obtidos, otimizar o tempo e permitir a análise de inúmeras variáveis simultaneamente (KRAMER, 1998; KOWALSKI, 1975). Suas aplicações envolvem: o planejamento e otimização de experimentos; a análise discriminante multivariada, ou análise exploratória, para reconhecimento de padrões; e a calibração multivariada, ou método de regressão, que permite a realização de análises quantitativas na presença de interferentes (SHI; YU, 2017; NETO; SCARMINIO; BRUNS, 2006).

As análises exploratórias mais utilizadas são a análise de componentes principais (PCA, do inglês *Principal Component Analysis*), análise de agrupamento hierárquico (HCA, do inglês *Hierarchical Cluster Analysis*) e, na aplicação do método de regressão, são comumente utilizadas a regressão de componentes principais (PCR, do inglês *Principal Components Regression*), regressão por quadrados mínimos parciais (PLSR, do inglês *Partial Least Squares Regression*) (NETO; SCARMINIO; BRUNS, 2006) e regressão múltipla linear (MLR, do inglês *Multiple Linear Regression*). A MLR visa modelar a relação entre uma variável dependente e duas ou mais variáveis independentes por meio de uma equação que descreva a

relação linear entre essas variáveis, permitindo fazer previsões ou inferências sobre a variável dependente (WANG *et al.*, 2020; FERREIRA, 2015). Essa técnica pode apresentar problemas com a colinearidade entre as amostras quando se tem uma amostragem pequena e, por isso, não foi considerada para esse estudo.

3.8.1 Análise de Componentes Principais

A análise de componentes principais, ACP ou do inglês PCA, é uma técnica multivariada, qualitativa, que faz combinações das variáveis originais, geralmente correlacionadas, e as transforma linearmente em novas variáveis, não correlacionadas, denominadas de componentes principais. Deste modo permite a redução do número de variáveis minimizando a perda de informações. Cada componente é uma combinação linear das variáveis originais e a essa está associado autovetores ortogonais, que se dispõe no espaço conforme a variância dos dados (BONA; MARÇO; VALDERRAMA, 2018).

As componentes principais são ordenadas pela fração de variabilidade do conjunto de dados, denominada de autovalor, em ordem decrescente, assim a primeira componente é aquela que possui a maior quantidade de variação no conjunto total em relação aos seguintes. Matematicamente, o algoritmo efetua a decomposição dos dados (matriz X) em duas matrizes menores com as projeções ortogonais das amostras no eixo da componente principal (PC, do inglês *Principal Component*) e a contribuição de cada variável original em relação a sua respectiva PC, além de considerar os resíduos que contém a variância não explicada pelo modelo (Equação 1) (BRERETON, 2018).

$$X = TP^t + E \quad (1)$$

Sendo T as projeções ortogonais das amostras no eixo de cada PC, P contribuição de cada variável e E os resíduos.

3.8.2 Análise de Agrupamento Hierárquico

Quando se tem um grande número de amostras é interessante buscar padrões para reuni-las em grupos mutuamente excludentes, baseados nas similaridades entre elas, de maneira que as amostras que possuem maior grau de associação fiquem agrupadas, criando uma homogeneidade no grupo, e que este possa se diferenciar

de outros, facilitando a interpretação dos resultados (FERREIRA, 2015; CORREIA *et al.*, 2007). Para isso é utilizada a Análise de Agrupamento Hierárquico, HCA, ou *Hierarchical Cluster Analysis*.

Esse método utiliza-se de uma análise exploratória das variáveis para encontrar aquelas responsáveis por agrupar semelhantes e maximizar a distância entre as diferenças (STATSOFT, 2004). Para estabelecer a distância entre as dissimilaridades existem algumas medidas matemáticas que determinam o quanto uma amostra é semelhante a outra, sendo as mais comuns a distância de Manhattan, de Mahalanobis e a Euclidiana (FERREIRA, 2015; BERRUETA *et al.*, 2007; MANLY, 2004). A distância Euclidiana é a mais utilizada na área da química e se baseia na norma da diferença vetorial para um espaço multidimensional, sendo a raiz quadrada do somatório dos quadrados das diferenças entre os valores de dois indivíduos (k e l) para todas as variáveis do sistema (Equação 2).

$$d_{kl} = \sqrt{[\sum_{i=1}^n (x_{ki} - x_{li})^2]} \quad (2)$$

Sendo d_{kl} a distância euclidiana e x_{ki} e x_{li} vetores das amostras k e l , na dimensão i , respectivamente.

Para realizar os agrupamentos hierárquicos, um dos métodos utilizados no HCA é o método de Ward. O método de Ward se baseia na soma dos quadrados das diferenças entre os valores de cada variável dentro de cada grupo. Ele busca combinar grupos de forma que a soma das variáveis dentro de cada grupo seja minimizada. O método de Ward é um método aglomerativo, o que significa que ele começa com cada observação como um grupo separado e, em seguida, funde os grupos sucessivamente para criar uma hierarquia de agrupamentos. Essa hierarquia é representada graficamente por meio de um dendrograma (GRANATO *et al.*, 2017).

O dendrograma é um diagrama bidimensional que agrupa dados por níveis de similaridade. Esse diagrama contribui para uma análise classificatória e comparativa dos dados (GRANATO *et al.*, 2017; CORREIA, 2018). O critério para o agrupamento das amostras em um grupo pode ser baseado na distância mais próxima, mais distante ou na distância média entre elas (BRERETON, 2018).

3.8.3 Regressão de Componentes Principais

A Regressão de Componentes Principais (PCR), é um método para relacionar a variância de uma variável de resposta (variável Y) à variância de vários preditores

(variáveis X), com fins explicativos ou preditivos. Ela combina os princípios da PCA e da regressão linear, permitindo lidar com conjuntos de dados multivariados de alta dimensionalidade (HOTELLING, 1993; JACKSON, 1991).

Na PCR, a PCA é aplicada inicialmente aos dados independentes para identificar as PCs mais significativas. Em seguida, as variáveis dependentes são regredidas nas componentes principais, obtendo-se os coeficientes de regressão para cada componente. Esses coeficientes são usados para construir um modelo de regressão linear que relaciona as componentes principais selecionadas à variável dependente (Equação 3). Um dos métodos para obtenção das PCs é dado com o algoritmo de Decomposição de Valor Singular (SVD, do inglês *Singular Value Decomposition*) (FERREIRA, 2015; STATSOFT, 2004).

$$Y_{pred} = Offset + \sum_{i=1}^n (PC_n \times Loadings_n) \quad (3)$$

Sendo Y_{pred} o valor previsto da variável resposta (variável dependente), *Offset* o deslocamento da reta ajustada, PC_n as componentes principais obtidas a partir dos dados de comprimento de onda das amostras, $Loadings_n$ os coeficientes do modelo PCR associados a cada PC.

Uma das principais vantagens da PCR é a capacidade de lidar com colinearidade, ou seja, quando as variáveis independentes estão altamente correlacionadas entre si. Ao reduzir a dimensionalidade dos dados, ela ajuda a evitar problemas associados à multicolinearidade e melhora a interpretação dos resultados (ESBENSEN, 2002; MARDIA; KENT; BIBBY, 1979).

3.8.4 Regressão de Mínimos Quadrados Parciais

A Regressão de Mínimos Quadrados Parciais (PLSR, do inglês *Partial Least Squares Regression*) é um modelo de regressão que modela as matrizes X e Y simultaneamente para encontrar variáveis latentes (ou ocultas) em X, que melhor irá prever as variáveis latentes em Y (Equação 4) (FERREIRA, 2015; STATSOFT, 2004).

$$Y_{pred} = Offset + \sum_{i=1}^n (PC_n \times Weight_n \times X_n) \quad (4)$$

em que, $Weight_n$ são os pesos associados a cada PC, aprendidos durante o processo de treinamento do modelo, e X_n os valores das amostras nas PCs.

Os componentes da PLS são semelhantes aos PCs, mas nesse caso são referidos como Fatores, eles são combinações lineares das variáveis originais e são obtidos por meio de um processo iterativo. O algoritmo de PLSR extrai

sequencialmente os componentes, maximizando a covariância entre as variáveis preditoras e de resposta, ao mesmo tempo em que maximiza a variância explicada pelos próprios componentes. Uma das vantagens da PLSR é sua capacidade de lidar com situações de multicolinearidade, em que as variáveis preditoras são altamente correlacionadas, além disso, fornece coeficientes de regressão para cada variável preditora, que podem ser usados para interpretar a força e a direção das relações entre as variáveis preditoras e de resposta. Esses coeficientes indicam quanto uma mudança unitária em uma variável preditora influencia a variável de resposta (FERREIRA, 2015; BERRUETA *et al.*, 2007).

Para avaliação do modelo construído a partir da PLSR, a medida mais comum é o coeficiente de determinação (R^2), que indica a proporção da variância nas variáveis de resposta explicada pelo modelo. Um valor de R^2 mais alto sugere um ajuste melhor do modelo aos dados. O algoritmo para o modelo de regressão utilizado em casos de grande número de variáveis e poucas amostras, como caso dessa pesquisa, é o Wide Kernel, uma variante do Kernel PLS baseada em Rannar *et al.* (1994), que não lida com valores omissos (SUN; ZHU; LIU, 2019).

4 METODOLOGIA

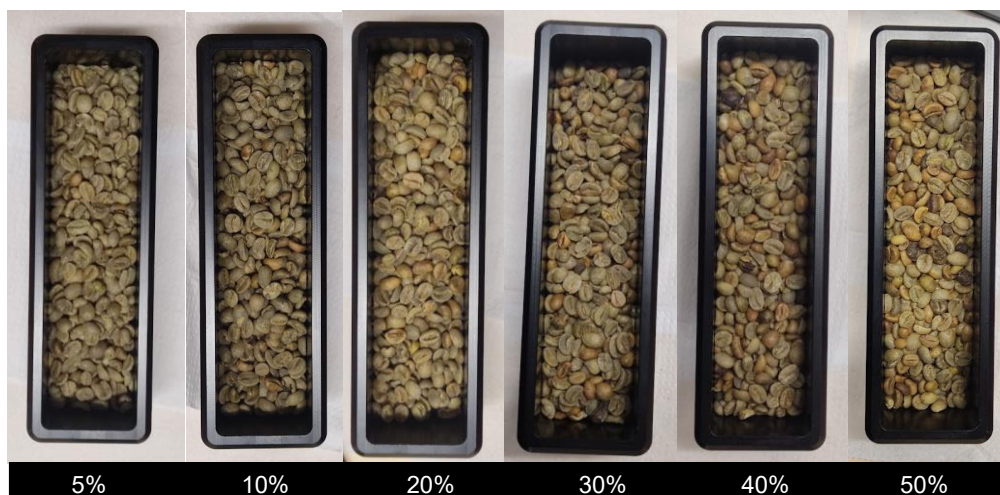
4.1 AMOSTRAS

As amostras de grãos de café cru beneficiados utilizadas foram disponibilizadas pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA), referentes às safras de 2020/21, as quais foram coletadas em diferentes regiões cafeeiras (BA, ES, MG, PR, RO, SP) e das espécies *Coffea arabica* (arábica) e *Coffea canephora* (canéfora), garantindo a representatividade amostral dos cafés produzidos no Brasil. Ao todo, foram analisadas 66 amostras, das quais 38 são de café arábica, representando 58% da amostragem, e 28 são de café canéfora, representando 42%. No Anexo A encontram-se informações referentes às amostras utilizadas.

4.2 PREPARAÇÃO DOS *BLENDS*

Os *blends* de café foram preparados em diferentes proporções de café arábica e café canéfora, sendo elas 5, 10, 20, 30, 40 e 50% (m/m) em relação à canéfora (Figura 8), para um total de 50 g do *blend*.

Figura 8. *Blends* preparados com café arábica e canéfora nas proporções de 5, 10, 20, 30, 40 e 50% (m/m) em relação à canéfora



Fonte: Própria autora (2023)

As amostras selecionadas para compor o *blend* foram SP099 para o arábica, que possui 43 defeitos, e RO006 para o canéfora, com 24 defeitos. A seleção dessas

amostras seguiu o critério de menor número de defeitos na classificação do grão, seguindo a COB, para minimizar influências advindas dessa variável. Os códigos utilizados para identificar as proporções dos *blends* nas amostras analisadas, bem como sua classificação quanto aos defeitos estão apresentados na Tabela 5 e 6.

Tabela 5. Composição dos *blends* de café arábica e canéfora utilizadas no estudo

Código	Proporções das misturas			
	Arábica ¹		Canéfora ²	
	Porcentagem (%) ³	Massa (g) ³	Porcentagem (%) ³	Massa (g) ³
C_100	0	0	100	50
C_50	50	25	50	25
C_40	60	30	40	20
C_30	70	35	30	15
C_20	80	40	20	10
C_10	90	45	10	5
C_5	95	47,5	5	2,5
C_0	100	50	0	0

¹amostra SP099, ²amostra RO006

³valores aproximados uma vez que, por se tratar de grãos, não foi possível a pesagem exata

Fonte: Própria autora (2023)

Tabela 6. Classificação manual feita por classificador do MAPA seguindo a COB quanto ao número de defeitos em 300 g para os cafés arábica (SP099) e canéfora (RO006) utilizado para o preparo dos *blends*

Tipo do defeito	Café arábica			Café canéfora		
	Massa (g)	Nº grãos	Defeitos	Massa (g)	Nº grãos	Defeitos
Preto	0,03	1	1,0	0,00	0,0	0,0
Ardido	1,16	14	7,0	2,40	26	13,0
Brocado Sujo	0,60	5	2,5	0,00	0	0,0
Verde	3,51	35	7,0	0,00	0	0,0
Casca média/pequena	0,00	0	0,0	0,11	2	0,8
Pau médio	0,00	0	0,0	0,06	2	4,0
Quebrado	11,22	125	25,0	3,55	33	6,6
Total	17	180	43	6	63	24
Tipo	4-45			3-45		

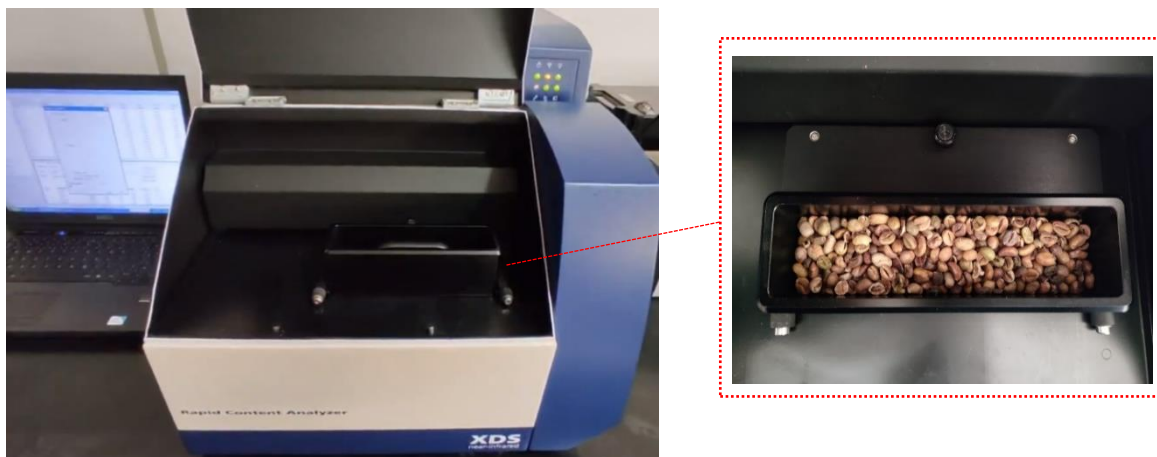
Fonte: Própria autora (2023)

4.3 AQUISIÇÃO DOS ESPECTROS DE INFRAVERMELHO PRÓXIMO

Para a aquisição dos espectros NIRS foram pesados 50 g de grãos crus de

cada amostra e introduzidas à uma cela do tipo “carrinho” (Figura 9), procedendo-se a leitura para obtenção dos dados espectrais por meio do espectrômetro NIRS Foss® (XDS *Rapid Analyser Content*, Dinamarca). O “carrinho”, utilizado no equipamento de espectrômetro de infravermelho para análise de matérias sólidas do tipo grãos, foi selecionado para o modo *moving* para garantir a incidência de luz ao longo de toda amostragem. As análises foram realizadas em duplicata, na faixa espectral do infravermelho próximo (750 a 2500 nm), com intervalo de 2 nm, e utilizada a média das leituras para os tratamentos, resultando numa matriz de 66x875, na qual as variáveis dependentes (linhas) correspondem as amostras de grãos de café cru e as variáveis independentes (colunas) correspondem a variação do comprimento de onda (FERREIRA, 2015). Para os *blends*, 16 leituras foram realizadas, duas para cada ponto de variação, gerando uma matriz de 8x875. Os dados espectrais foram registrados como absorbância ($\log(1/R)$, onde R é a refletância) e tratados por meio de análises quimiométricas.

Figura 9. Cela do tipo “carrinho” utilizada no espectrômetro de infravermelho para as amostras de grãos de café



Fonte: Própria autora (2023)

4.4 ANÁLISES ESTATÍSTICAS

Para reduzir fontes de variações irrelevantes aos objetivos, foram aplicados os pré-processamentos de correção do espalhamento multiplicativo (MSC) e filtro de suavização Savitzky-Golay com primeira derivada, janela de 9 pontos e polinômio de segunda ordem, e segunda derivada, janela de 7 pontos e polinômio de segunda ordem.

Para a PCA e PCR, o algoritmo utilizado foi o SVD, enquanto para a PLSR utilizou-se o algoritmo Wide Kernel. Os dois modelos utilizados nos *blends*, PCR e PLSR, foram validados pelo método LOO com os dados centrados na média. Na análise de HCA, realizada exclusivamente para as amostras puras, foi utilizado o método de Ward para aglomeração com corte esquemático na maior distância Euclidiana.

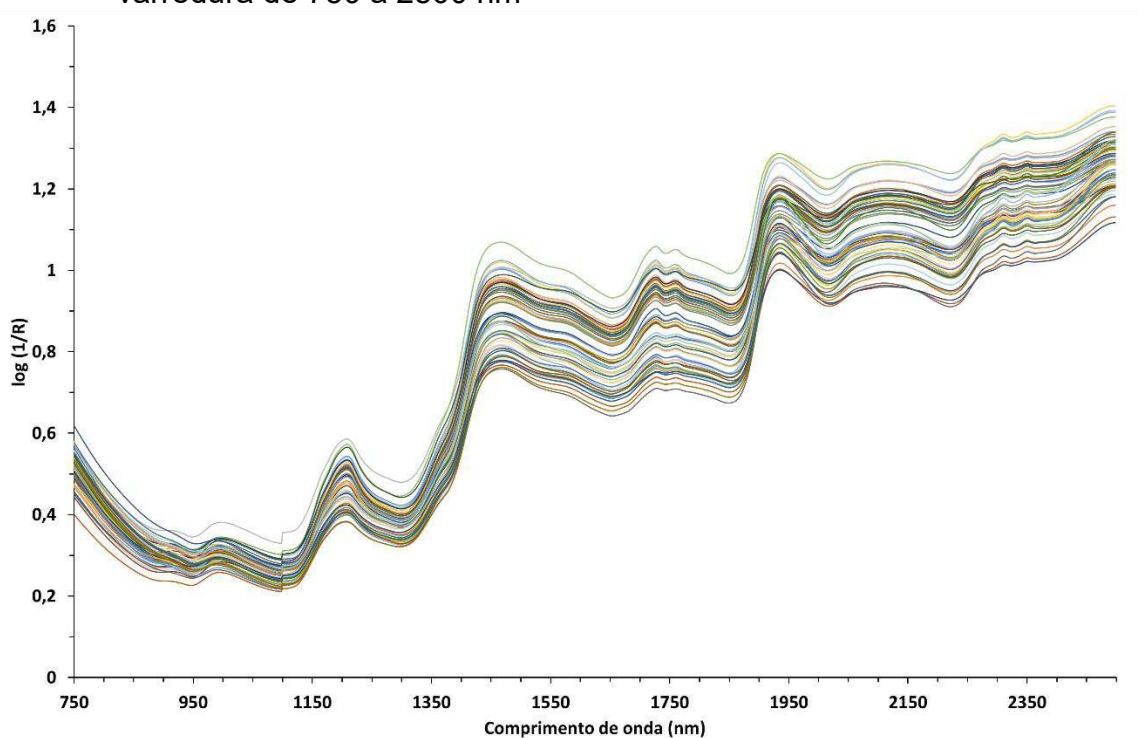
Os tratamentos estatísticos foram realizados utilizando o *software* The Unscrambler® X (Camo, A/S, Oslo, Noruega), versão 10.4 e o Excel® (Microsoft Corporation, Estados Unidos), versão 2013.

5 RESULTADOS E DISCUSSÃO

5.1 ESPECTROS DE INFRAVERMELHO PRÓXIMO

Os espectros obtidos no NIRS para as amostras de café cru ($n = 66$), sem pré-processamento (dados brutos), estão apresentados na Figura 10.

Figura 10. Sobreposição da média da duplicata dos espectros NIRS para 66 amostras de grãos de cafés crus, sendo 38 de arábica e 28 de canéfora, com varredura de 750 a 2500 nm

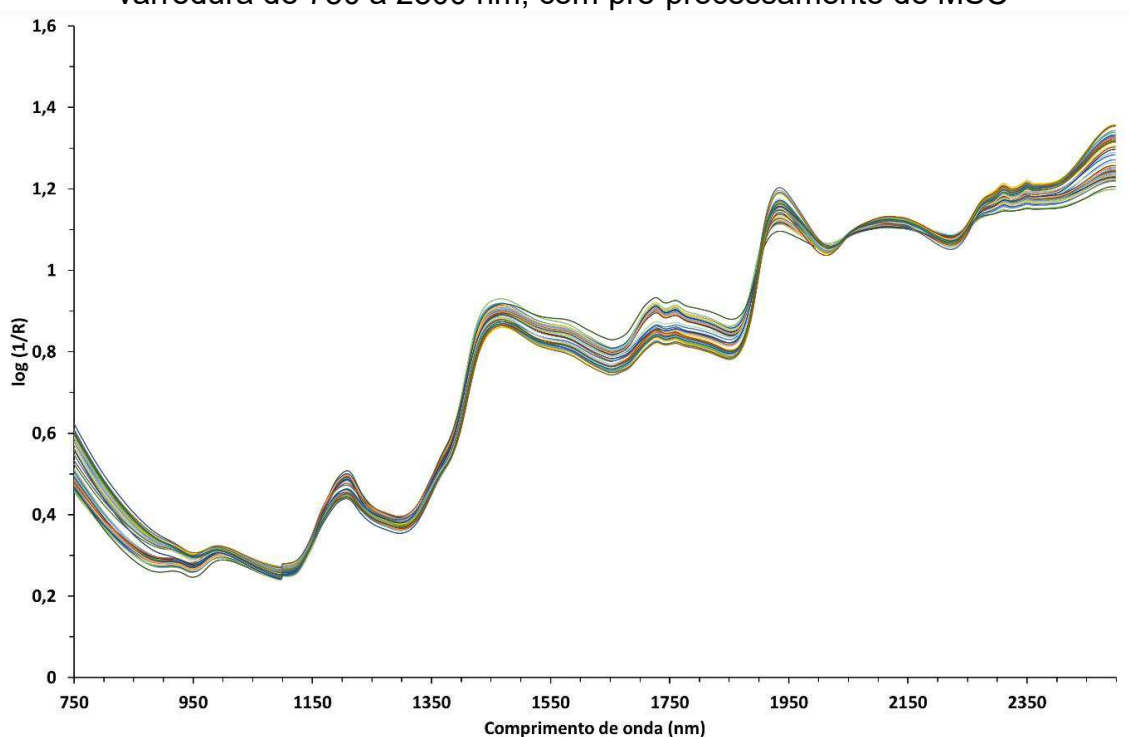


Fonte: Própria autora (2023)

Como as amostras analisadas são grãos de café cru beneficiados, vindos diretamente de cooperativas, sem terem sido passados em peneiras, esses grãos possuem irregularidades em seu tamanho que influenciam os espectros do infravermelho (FERREIRA, 2015). A utilização de amostras sólidas no NIRS tem como vantagem a minimização do preparo de amostras, mas, em contrapartida, a dispersão de luz causada pela granulometria é responsável por perturbações no espectro como deslocamentos da linha de base, mudanças de inclinação e perda da linearidade que, se não forem minimizadas, podem comprometer a confiabilidade dos modelos preditivos (BARBIN *et al.*, 2014; ESTEBAN-DÍEZ; GONZÁLEZ-SÁIS; PIZARRO, 2004).

Segundo Rinnan *et al.* (2009), existem vários métodos de pré-processamentos para corrigir a influência desses fenômenos físicos nos espectros. Assim, para corrigir o espalhamento de luz causado pela diferença na granulometria das amostras e rugosidades em sua superfície, foi aplicado o pré-processamento MSC, que consiste em uma transformação utilizada para corrigir os efeitos de espalhamento aditivos e multiplicativos na absorbância (FERREIRA, 2015). Os espectros com pré-processamento MSC podem ser visualizados na Figura 11.

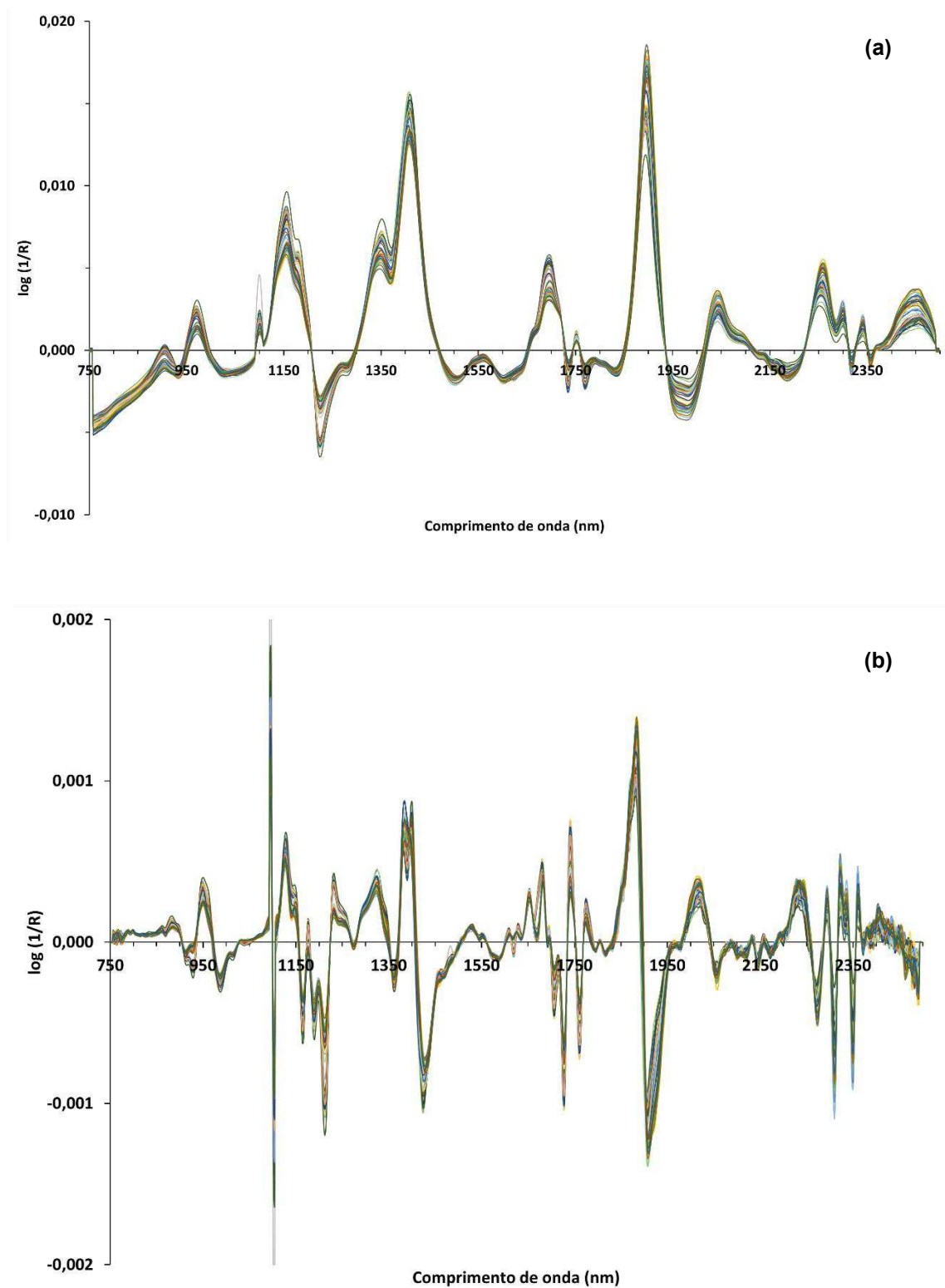
Figura 11. Sobreposição da média da duplicata dos espectros NIRS para 66 amostras de grãos de cafés crus, sendo 38 de arábica e 28 de canéfora, com varredura de 750 a 2500 nm, com pré-processamento de MSC



Fonte: Própria autora (2023)

Dando prosseguimento aos pré-processamentos, foi aplicado o filtro de suavização de Savitzky-Golay seguido de: primeira derivada (Figura 12a), com janela de 9 pontos e polinômio de segunda ordem; segunda derivada (Figura 12b), com janela de 7 pontos e polinômio de segunda ordem. A primeira derivada é utilizada para corrigir o deslocamento da linha de base, enquanto a segunda derivada corrige a variação sistemática observada pela inclinação da mesma (MOGHIMI *et al.*, 2010; SAVITZKY; GOLAY, 1964). Aos espectros das derivadas, também foi realizado o processamento de MSC.

Figura 12. Filtro Savitzky-Golay com derivadas e pré-processamento de MSC. (a) primeira derivada com janela de 9 pontos e polinômio de segunda ordem e (b) segunda derivada com janela de 7 pontos e polinômio de segunda ordem



Fonte: Própria autora (2023).

É possível observar nos espectros da Figura 11 bandas intensas que podem

ser atribuídas as funções orgânicas presentes na matriz e, em conjunto com a derivada segunda (Figura 12b), foram identificados com maior precisão seus comprimentos de onda por meio dos pontos de inflexões e curvaturas. A esses comprimentos de onda foram atribuídos os possíveis componentes que, segundo a Figura 7 e demais autores, absorvem naquela região do infravermelho.

Próxima à faixa espectral do visível, em torno de 750 até 1050, não foram encontrados muitos estudos de componentes do café que são referentes a essa região, além de que o espectro não apresentou bandas que se sobressaíssem. Já entre 1110 e 1224 nm, sinais intensos foram identificados, essa faixa geralmente está relacionada ao segundo sobretom –CH dos componentes principais do café, tais como cafeína, ácidos clorogênicos e carboidratos. O ponto de inflexão entorno de 1900 nm, região referente a bandas de combinação e sobretons, pode estar associada ao grupo funcional amida (CONH₂) presente na cafeína, nessa região é característica a vibração do segundo sobretom de estiramento de carbonilas (C=O) (FERREIRA, 2015 *apud* XIAOBO *et al.*, 2010). Acima de 2100 nm apresenta uma região mais complexa, com combinações de funções orgânicas absorvendo energia, observado no espectro pela quantidade de bandas com máximos e mínimos bem definidos. Essa faixa espectral está atribuída a compostos como cafeína, lipídios e ácidos clorogênicos (MARQUETTI *et al.*, 2016).

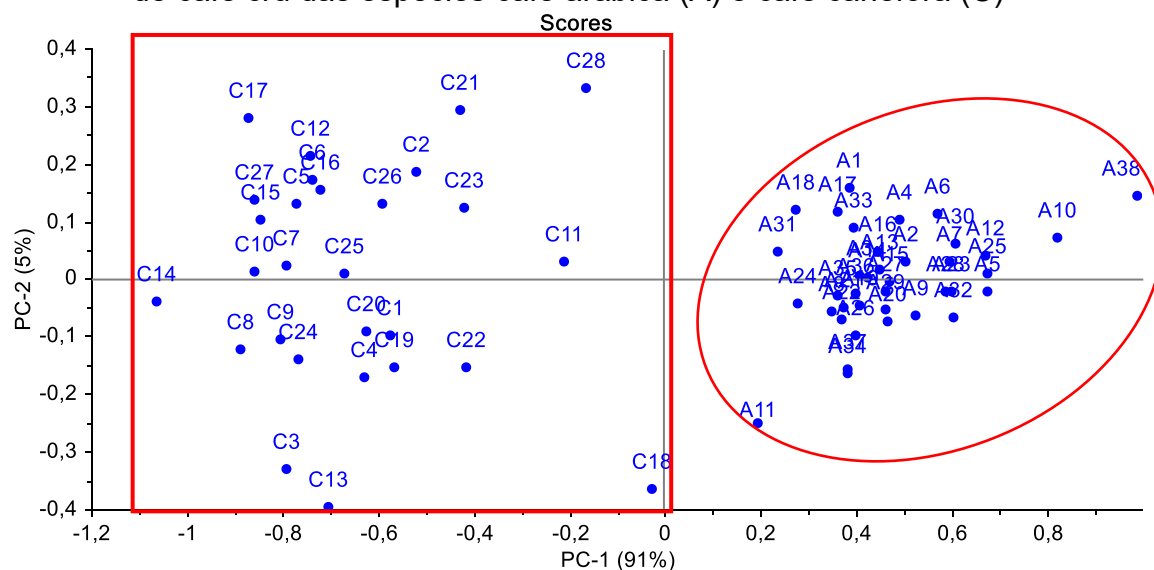
5.2 ANÁLISES QUALITATIVAS

As análises qualitativas realizadas foram a PCA e a HCA, ambas com os dados tratados com pré-processamento de MSC. Os tratamentos com o filtro de Savitzky-Golay e derivadas foram aplicados, mas não obtiveram boas relações dos vetores, dificultando o entendimento das relações entre os comprimentos de onda do NIRS e as amostras.

5.2.1 ANÁLISE DE COMPONENTES PRINCIPAIS

O gráfico de *scores*, em que são mostradas a primeira componente principal (PC-1) e a segunda componente principal (PC-2), que juntas explicam 96% da variância total dos dados, está apresentado na Figura 13.

Figura 13. Gráfico de *scores*, no plano bidimensional das componentes principais 1 e 2, referentes aos espectros na região do NIRS das 66 amostras de grãos de café cru das espécies café arábica (A) e café canéfora (C)

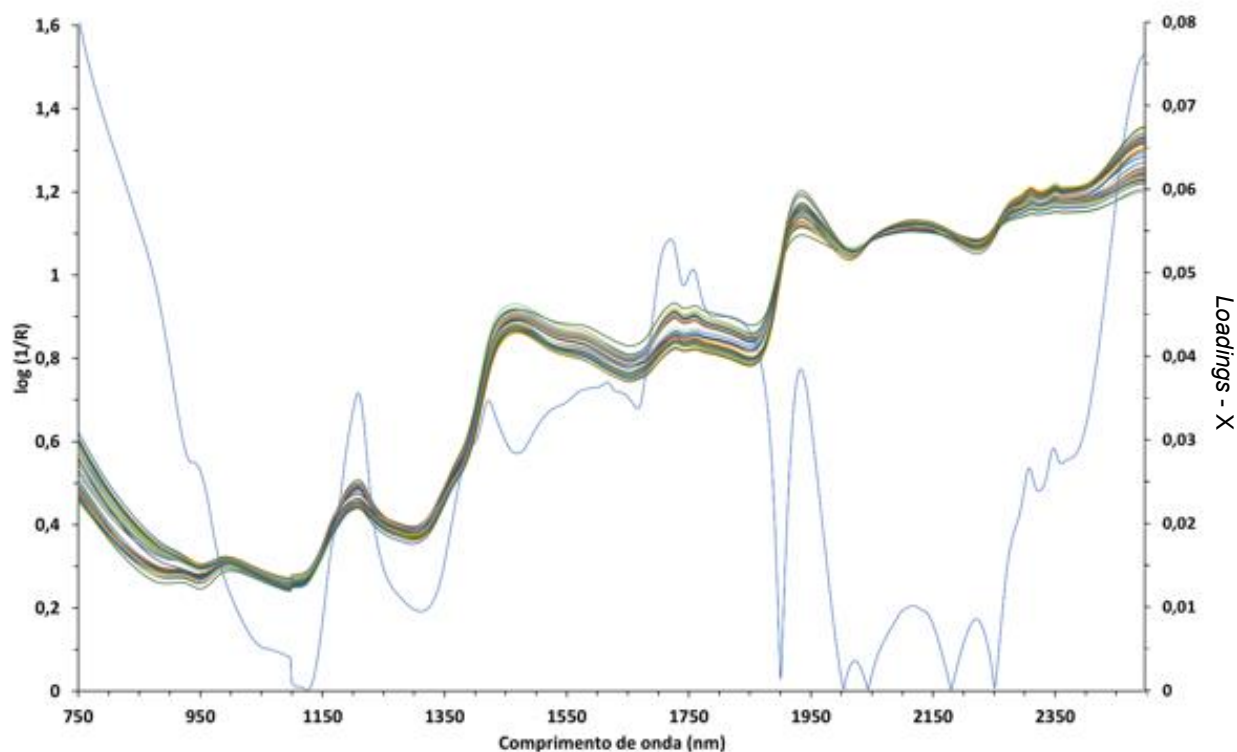


Fonte: Própria autora (2023)

Cada ponto do gráfico representa uma amostra individual de café. As amostras mais próximas, em aglomerados, possuem maiores semelhanças em termos de suas características de comprimento de onda, já as mais distantes no gráfico são mais dissimilares. Observando a Figura 13, nota-se a formação de dois grandes grupos que foram separados em relação as espécies do café – canéfora (C) e arábica (A). O PC-1, por si só, é altamente discriminante, afirmando que os espectros NIR das duas espécies possuem diferenças suficientemente significativas que possibilitam suas discriminações.

Uma vez percebida a boa separação por espécies com o espectro em sua totalidade, foi estudada suas regiões separadamente com o intuito de explorar as características químicas de cada grupo de comprimento de onda, que são importantes para distinguir as amostras de café. Essas informações podem ser relevantes para controle de qualidade e identificação dos grãos puros. Para realizar as separações dos comprimentos de onda dos espectros, primeiramente é necessário compreender as regiões que possuem maior contribuição na discriminação das espécies, para isso foi analisado o gráfico de *loadings-X* (Figura 14).

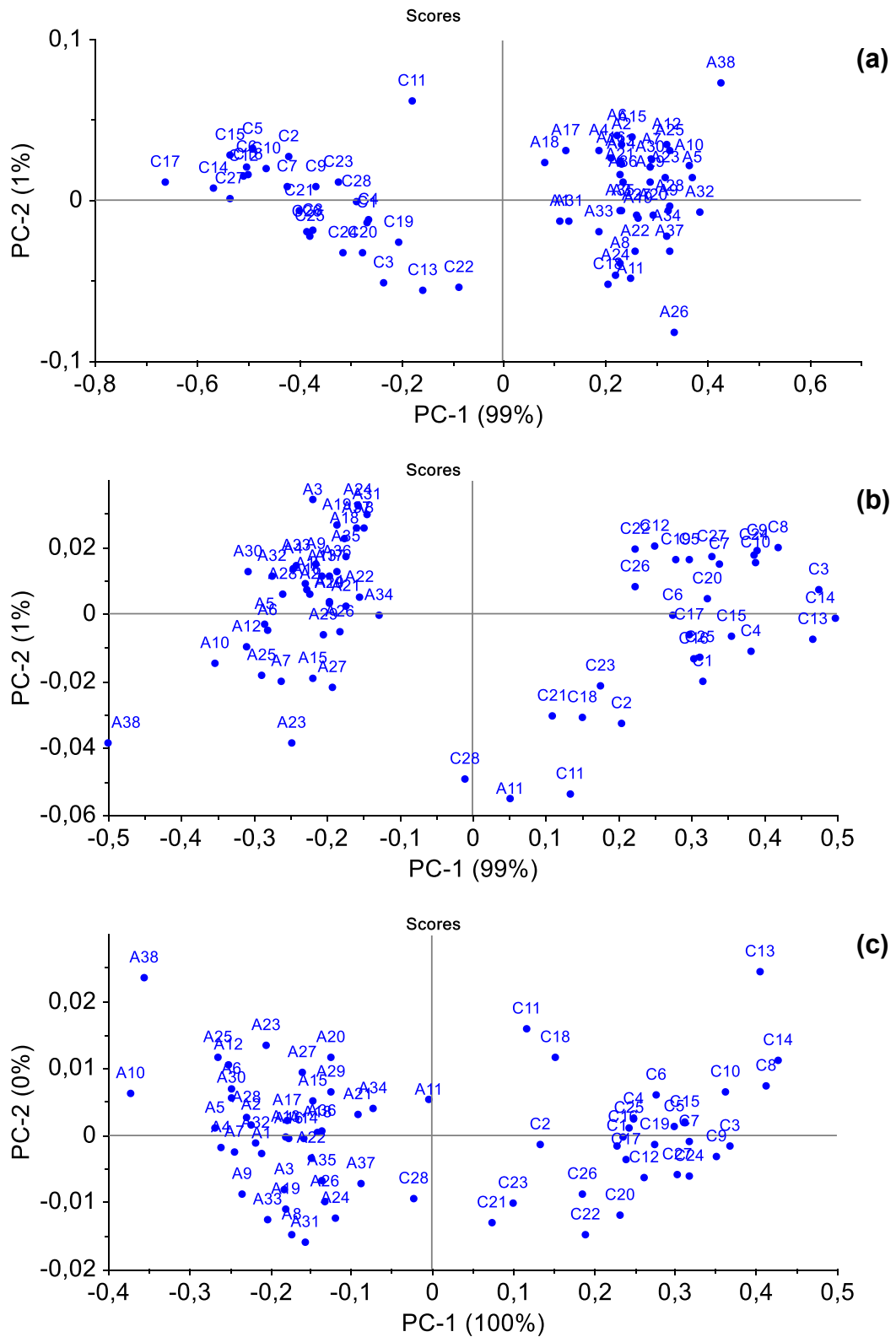
Figura 14. Conjunto de espectros das 66 amostras de café e gráfico de *loadings-X* (linha azul) para a PC-1, com as variáveis do eixo X sendo os comprimentos de onda de 750 a 2500 nm



Fonte: Própria autora (2023)

Os *loadings-X* representam as regiões com maior correlação com a componente principal e que melhor as explicam. Os valores absolutos de *loadings-X*, considerando a PC-1, que são mais significativos, coincidem com as regiões do espectro NIR que se sobressaem, podendo ser indicadores importantes na identificação e distinção entre as espécies. Por meio desse gráfico foi determinada as seguintes regiões para separação do espectro: 750 a 900; 1680 a 1870 e 2420 a 2500 nm. Essa escolha se baseou no fato de serem as regiões com maiores valores absolutos de *loadings-X*, acima de 0,04. Os gráficos de *scores* para cada região são apresentados a seguir, na Figura 15.

Figura 15. Gráfico de *scores*, no plano bidimensional da PC-1 e PC-2, referentes aos comprimentos de onda: (a) 750-900 nm; (b) 1680-1700 nm; (c) 1700-2000 nm e (d) 2000-2500 nm, do espectro NIRS



Fonte: Própria autora (2023)

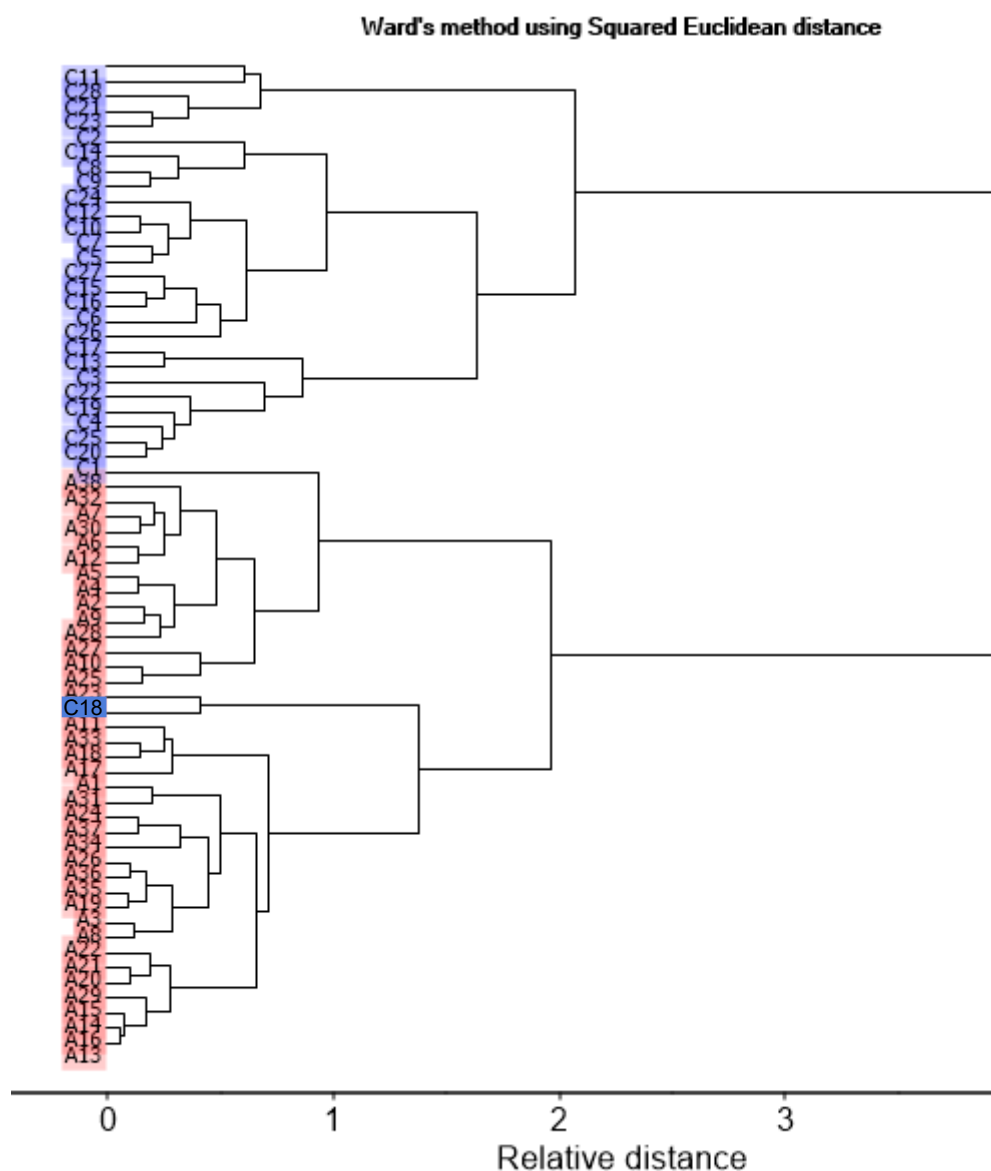
Analisando os gráficos de *scores*, nota-se que a espécie arábica tende a possuir maior similaridade entre seus grãos quando comparados ao do canéfora, percebida pelo melhor agrupamento delas, indicando uma possibilidade de identificação para controle de qualidade. Com a fragmentação dos espectros nos comprimentos de onda de maior contribuição para a discriminação foi observado um bom grau de separação entre as espécies, mas, para algumas amostras, suas similaridades foram compatíveis com a outra espécie. Esse comportamento pode ser explicado pelo fato de que alguns de seus compostos não possuem grandes diferenças de concentração entre as espécies, por isso a importância de se trabalhar com o espectro em sua totalidade, que possibilita a melhor discriminação.

5.2.2 ANÁLISE DE AGRUPAMENTO HIERÁRQUICO

O dendrograma (Figura 16) foi obtido por meio da HCA pelo método de Ward, com corte esquemático no quadrado da distância Euclidiana, que confirma a separação resultante da PCA (Figura 13), onde são revelados os dois grandes grupos.

O corte esquemático no dendrograma se deu na distância relativa de 5, correspondendo à maior distância Euclidiana existente entre os grupos. Os grupos formados demonstraram similaridades entre si, apresentando coerência nos resultados, uma vez que se trata de espécies distintas, com exceção da amostra C18. A amostra codificada como C18 foi agrupada erroneamente no grupo que se refere a espécie arábica.

Figura 16. HCA referente às amostras de grãos de cafés crus das espécies café canéfora (C, em azul) e café arábica (A, em vermelho)



Fonte: Própria autora (2023)

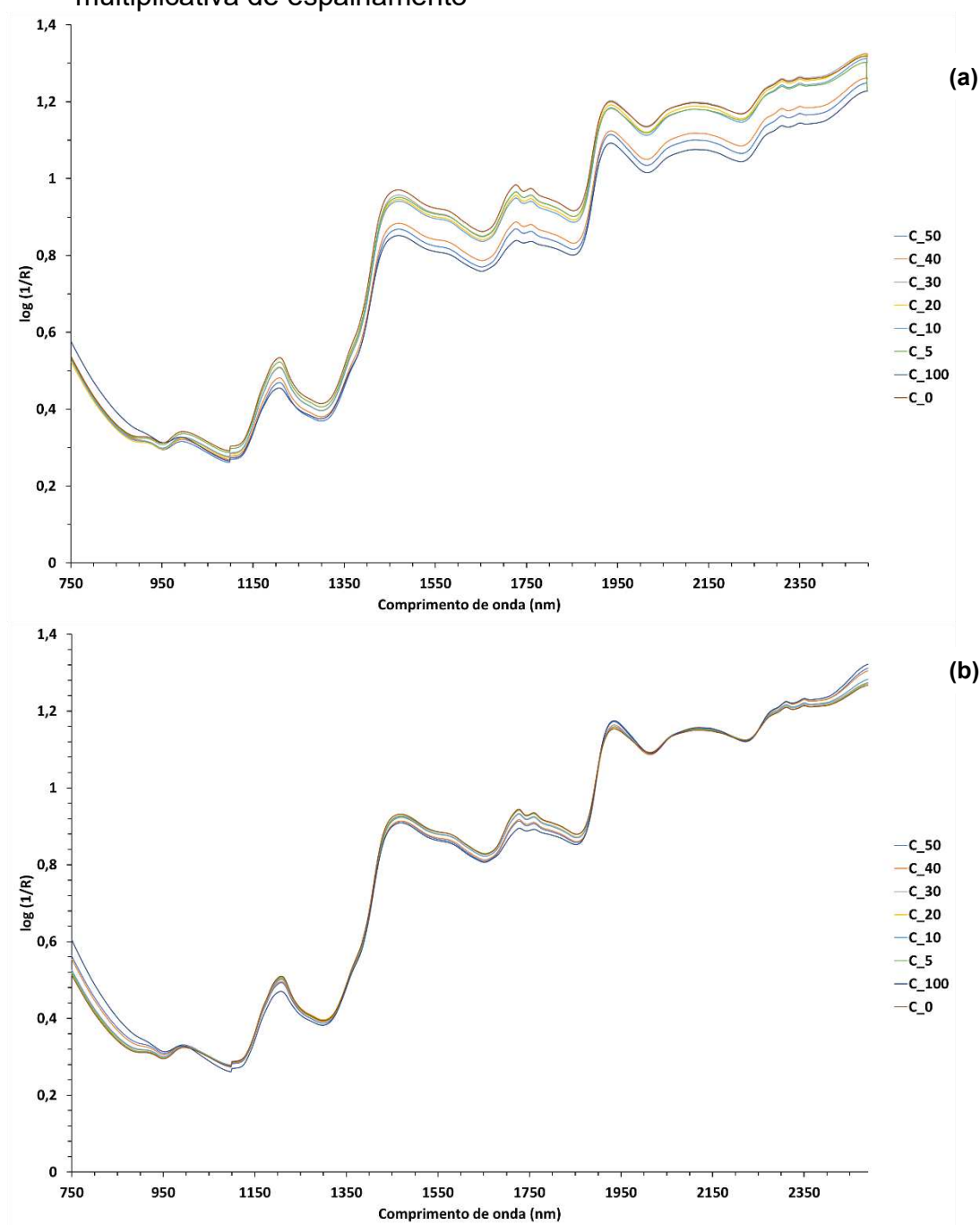
Os *clusters* menores, em alguns casos, foram aglomerados quanto a sua origem de plantio. Esse comportamento pode ser explicado pelo fato de mesmas espécies apresentarem diferenças em sua composição quando cultivadas em diferentes regiões, solos e adversidades climáticas (LAMBOT *et al.*, 2016; PERRONE *et al.*, 2008).

5.2.3 AVALIAÇÃO DO COMPORTAMENTO EM *BLENDS*

Uma vez avaliada a qualidade de separação das amostras puras, quando

estudada toda a extensão do espectro de infravermelho, foram investigados os mesmos parâmetros para os *blends*, a fim de avaliar seu comportamento em diferentes proporções. Os espectros brutos e tratados com MSC para os *blends* são apresentados na Figura 17.

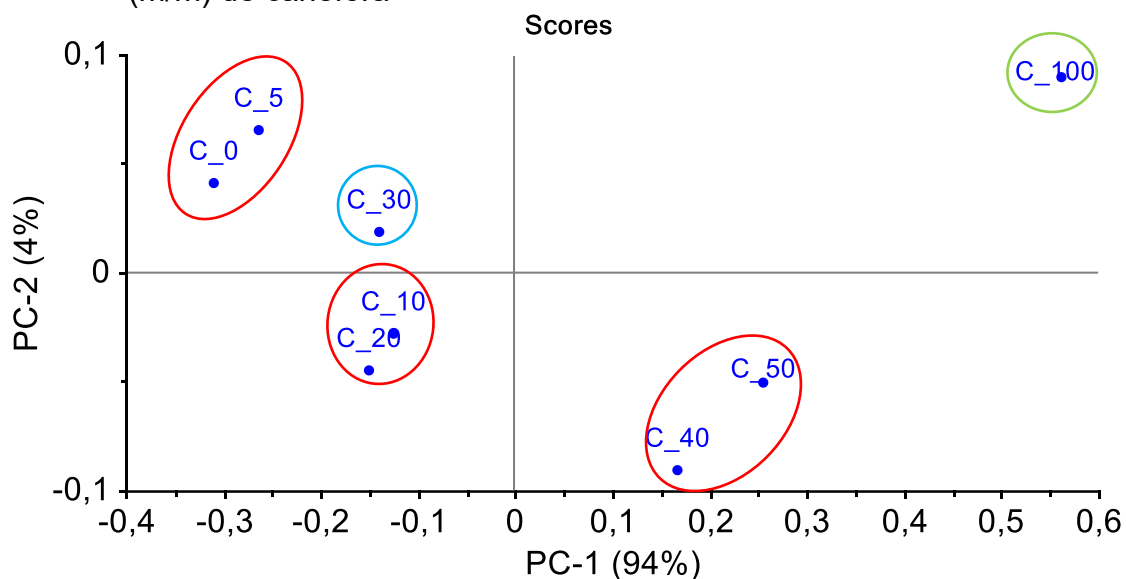
Figura 17. Sobreposição da média da duplicata dos espectros NIRS para 8 amostras de *blends* de grãos de cafés crus, com varredura de 750 a 2500 nm, para (a) espectro bruto e (b) espectro com pré-processamento de correção multiplicativa de espalhamento



Fonte: Própria autora (2023)

A partir dos dados espectrais, o método exploratório de PCA foi aplicado. O gráfico de *scores*, Figura 18, apresenta as componentes PC-1 e PC-2, que em conjunto explicam 98% da variância total dos dados.

Figura 18. Gráfico de *scores*, no plano bidimensional das componentes principais 1 e 2, referentes aos espectros na região do NIRS das 8 amostras de *blends* de grãos de cafés crus nas proporções 0, 5, 10, 20, 30, 40, 50 e 100% (m/m) de canéfora



Fonte: Própria autora (2023)

Observa-se que os *blends* apresentaram uma divisão com base na porcentagem de cada espécie, cuja codificação pode ser consultada na Tabela 5. A amostra pura de canéfora foi a que apresentou maior dissimilaridade entre as demais, resultado que indica a possibilidade de utilizar a técnica do NIRS para auxiliar na classificação da matéria-prima empregada nos cafés torrados e moídos comercializados. Também foi observado um adequado grau de distinção nas porcentagens mais baixas, como em 5% e nas proporções de 40 e 50%. As misturas nas porcentagens de 10, 20 e 30% de canéfora não apresentaram grandes diferenças discriminativas.

5.3 ANÁLISES QUANTITATIVAS

Visto que as duas espécies de café arábica e canéfora podem ser discriminadas por NIRS pela separação quando se utilizou toda a extensão do espectro de infravermelho, além de se observar a variabilidade dos *blends*, foram

desenvolvidos os modelos de regressão PCR e PLSR para estabelecer relações quantitativas entre as variáveis e estimar os valores numéricos para as amostras de *blends* de cafés em diferentes proporções.

5.3.1 REGRESSÃO POR COMPONENTES PRINCIPAIS

O método PCR foi aplicado para 6 componentes principais, número suficiente para capturar a maior parte da variabilidade dos dados. Os coeficientes de regressão das componentes principais, obtidos tanto para a calibração quanto para validação, se encontram na Tabela 7. Utilizou-se o método de validação cruzada LOO, por ser o melhor em situações em que a quantidade de amostras é insuficiente para separar em grupos de calibração e validação.

Tabela 7. Parâmetros estatísticos dos conjuntos de calibração e validação por LOO, obtidos a partir do modelo PCR, para 8 amostras de *blends* de canéfora no arábica de 0, 5, 10, 20, 30, 40, 50 e 100% de grãos de cafés crus

Componente	Calibração		Validação	
	RMSE	R ²	RMSE	R ²
PC1	7,983	0,931	13,547	0,848
PC2	5,958	0,962	10,669	0,906
PC3	5,949	0,962	11,539	0,890
PC4	3,872	0,983	11,028	0,899
PC5	3,844	0,984	10,763	0,904
PC6	3,618	0,986	14,761	0,820

RMSE: *Root Mean Square Error* (Erro Médio Quadrático)

Fonte: Própria autora (2023)

Observando os resultados dos parâmetros estatísticos obtidos, como esperado o modelo de calibração apresentou coeficientes de determinação ($R^2 > 0,93$) maiores que para o conjunto de validação ($R^2 > 0,82$), valores satisfatórios considerando o pequeno número de amostras que foram avaliadas. Embora os erros médios (RMSE) foram menores para o conjunto de calibração e foram sendo reduzidos à medida que se aumentou as PCs, esta tendência não foi observada para a validação. Tanto para calibração quanto para validação os valores dos erros quadráticos médios foram considerados altos, indicando que existem limitações para o modelo construído.

5.3.2 REGRESSÃO POR MÍNIMOS QUADRADOS PARCIAIS

O algoritmo utilizado no modelo PLSR foi o Wide Kernel, que nada mais é que uma variante do Kernel PLS que possui desempenho melhor para dados contendo um grande número de variáveis e relativamente poucas amostras (dados “short and fat”). A implementação é baseada em Ränner *et al.* 1994 e não lida com valores ausentes.

Na Tabela 8 são apresentados os resultados dos indicadores estatísticos obtidos na calibração e validação, igualmente utilizando LOO para validação, na construção do modelo de PLSR com 6 variáveis latentes. As variáveis latentes são úteis porque podem capturar relações complexas e não lineares entre as variáveis independentes e dependentes, e podem fornecer uma representação mais compacta dos dados. Além disso, as variáveis latentes podem ser interpretadas em termos das contribuições das variáveis originais em cada componente latente.

Tabela 8. Parâmetros estatísticos dos conjuntos de calibração e validação por LOO, obtidos a partir do modelo PLSR, para 8 amostras de *blends* de canéfora no arábica de 0, 5, 10, 20, 30, 40, 50 e 100% de grãos de cafés crus

Fator	Calibração		Validação	
	RMSE	R ²	RMSE	R ²
1	7,818	0,934	13,412	0,851
2	5,648	0,965	8,261	0,943
3	3,318	0,988	10,969	0,900
4	2,995	0,990	9,243	0,929
5	1,622	0,997	13,755	0,843
6	0,095	0,999	14,763	0,820

RMSE: *Root Mean Square Error* (Erro Médio Quadrático)

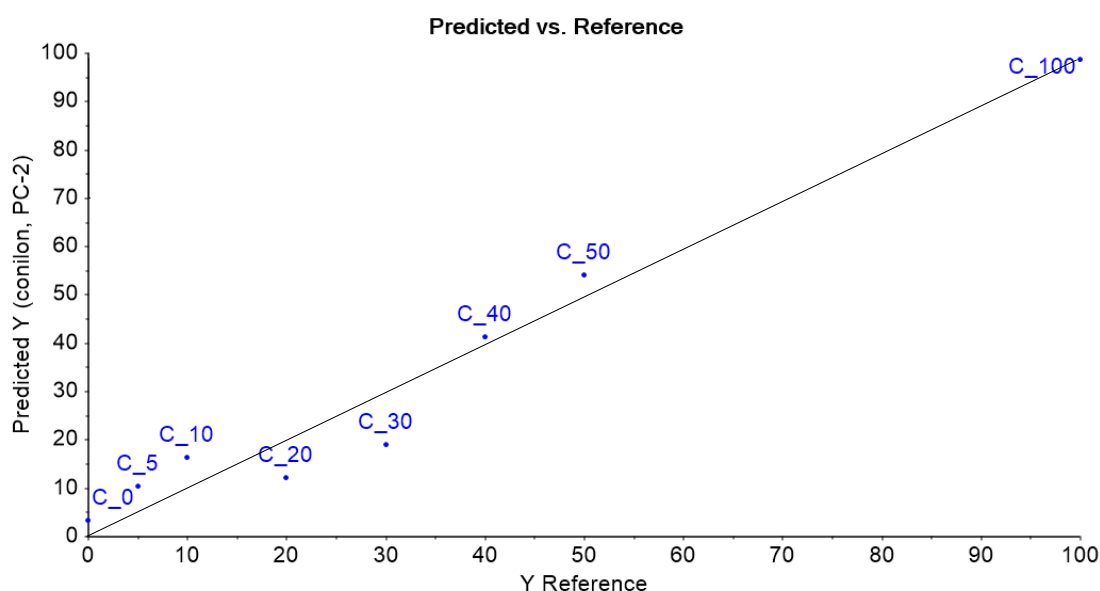
Fonte: Própria autora (2023)

Assim como os resultados obtidos pela PCR, a PLSR também apresentou valores de RMSE relativamente altos, mas aceitáveis considerando a matriz de análise, e com níveis menores de erro quando comparado com o primeiro modelo. O R² apresentou valores melhores que do modelo anterior e mais próximo de 1, indicando que esse modelo melhor se ajusta aos dados.

5.4 PREDIÇÃO DOS *BLENDS* A PARTIR DOS MODELOS

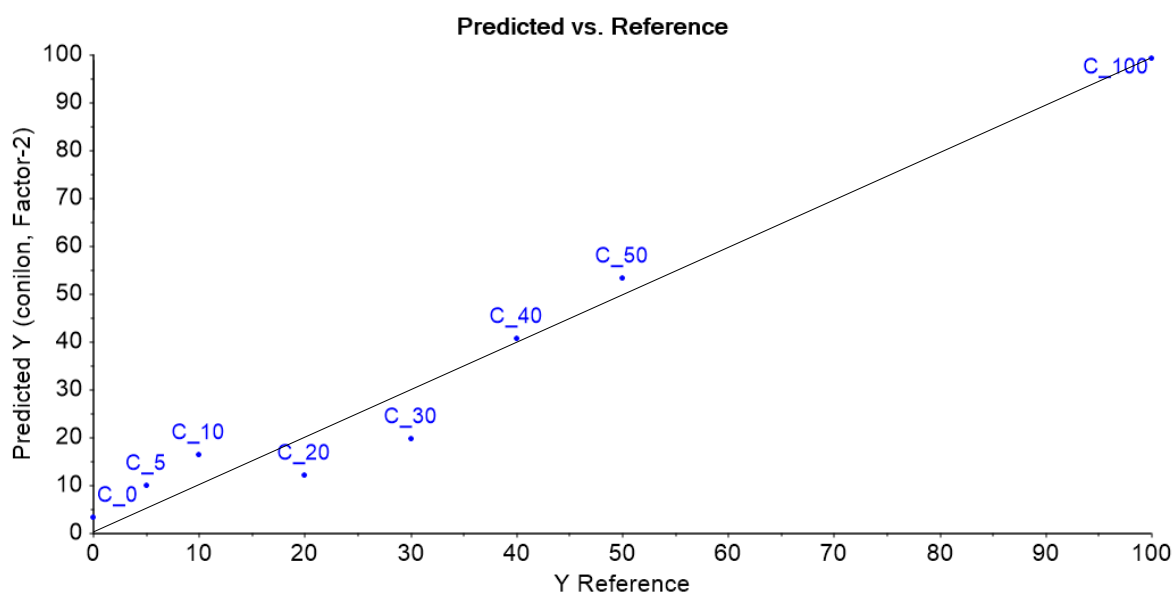
Tendo os modelos de predição construídos, PCR e PLSR, os dados obtidos no espectro para os *blends* foram avaliados por cada um dos dois, para verificarmos suas capacidades de predições nas diversas proporções, além de avaliar o modelo que melhor se ajusta para essas amostras. Baseando-se no menor erro quadrático RMSE para os conjuntos de validação, foi considerado duas componentes principais (PC-2) para predição dos modelos de PCR (Figura 19) e dois fatores (Fator-2) para PLSR (Figura 20).

Figura 19. Gráfico de valores Y referência versus Y previstos para o modelo PCR com PC-2, para 8 amostras de *blends* de grãos de café crus



Fonte: Própria autora (2023)

Figura 20. Gráfico de valores Y referência versus Y previstos para o modelo PLS com Fator-2, para 8 amostras de *blends* de grãos de café crus



Fonte: Própria autora (2023)

Analisando os gráficos num primeiro momento, observa-se que os modelos apresentaram uma predição similar, ambos tendendo a superestimar os valores de canéfora nos *blends* em baixas concentrações e subestimá-los nas proporções de 20 e 30% das misturas canéfora/arábica. Os resultados obtidos nos gráficos estão apresentados na Tabela 9, para melhor análise das predições.

Tabela 9. Valores preditos pelos modelos PCR e PLSR e seus desvios para 8 amostras de *blends* de grãos de café crus, nas proporções de 0, 5, 10, 20, 30, 40, 50 e 100% de canéfora em arábica.

Amostra (% canéfora)	MODELOS			
	PCR		PLSR	
	Predito	Erro	Predito	Erro
C_100	98,60	1,40	99,21	- 0,79
C_50	54,08	4,08	53,89	3,89
C_40	41,27	1,24	40,69	0,69
C_30	18,88	-11,12	19,76	-10,24
C_20	12,24	- 7,76	12,22	- 7,78
C_10	16,42	6,42	16,37	6,37
C_5	10,31	5,31	10,11	5,11
C_0	3,21	3,21	3,25	3,25

Fonte: Própria autora (2023)

Os modelos apresentaram melhor concordância entre o referencial e o predito para os *blends* cujas proporções estão acima de 40%. O modelo PLSR apresentou erros menores com relação a PCR, indicando uma melhor predição, contudo, ambos obtiveram maiores erros para C_10, C_20 e C_30, confirmando que os modelos podem ser utilizados para predições de *blends* quando o objetivo é estimar misturas com proporções mais elevadas.

Ressalta-se que validações adicionais deverão ser feitas em trabalhos futuros com um conjunto extra de amostras para validação, visando assegurar que não ocorra possíveis vícios no conjunto de validação.

6 CONSIDERAÇÕES FINAIS

A espectroscopia de infravermelho próximo associada a quimiometria se mostrou uma análise de classificação rápida e direta para diferenciar os cafés beneficiados de grãos crus das espécies arábica e canéfora. A análise empregando PCA explicou 96% da variância dos dados pelas componentes 1 e 2 e a HCA separou em dois grandes grupos distintos baseados nas duas espécies, tendo uma amostra de canéfora sido agrupada ao arábica dentre as 66 amostras analisadas.

Na análise dos *blends* de 0, 5, 10, 20, 30, 40, 50 e 100% embora os erros quadráticos médios tenham sido elevados, os modelos preditivos PCR e PLSR foram capazes de estimar os valores das proporções de canéfora em arábica em relação ao referencial para *blends* acima de 40%. O modelo PLSR obteve os menores erros e maior R^2 , mostrando ser o modelo que melhor se aplica aos dados.

Ressalta-se que estes resultados mostram o potencial de aplicação da técnica, mas são necessárias análises de amostras adicionais, para garantir uma classificação mais robusta para rotina, assim como, uma ampliação do conjunto de validação para uma maior precisão do modelo.

O método desenvolvido mostrou-se como alternativa verde e rápida de classificação das espécies de cafés, com potencial de predição de suas proporções em misturas, que tende a uma maior robustez com o acréscimo da variabilidade amostral.

REFERÊNCIAS

ABIC, Associação Brasileira da Indústria do Café. A crise de 1929. Brasil, 2021. Disponível em: <https://www.abic.com.br/tudo-de-cafe/a-crise-de-1929/#:~:text=A%20crise%20de%201929%20afetou,e%20queimou%20toneladas%20de%20caf%C3%A9>. Acesso em: 07 mar. 2022.

ABIC, Associação Brasileira da Indústria do Café. Disponível em: <https://www.abic.com.br/>. Acesso em: 02 fev. 2022.

ABIC, Associação Brasileira da Indústria do Café. Tendências de consumo de café. Brasil. ABIC: Ivani Rossi, 2010.

ABRAHÃO, S.A.; PEREIRA, R.G.F.A.P.; LIMA, A.R. FERREIRA, E.B.; MALTA, M.R. Compostos bioativos em café integral e descafeinado e qualidade sensorial da bebida. **Pesquisa Agropecuária Brasileira**, v. 43, n. 12, p. 1799-1804, 2008.

ADNAN, A.; NAUMANN, M.; MÖRLEIN, D.; PAWELZIK, E. *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**, 9 (6), 788, 2020.

AGNOLETTI, B. Z.; OLIVEIRA, E. C. da S.; PINHEIRO, P. F.; SARIVA, S. H. Discriminação de café arábica e conilon utilizando propriedades físico-químicas aliadas à quimiometria. **Revista Virtual de Química**. v. 11, n. 3, p. 785-805. 2019.

AGUIAR, A. T. E. *et al.* Diversidade química de cafeeiros na espécie *Coffea canephora*. **Bragantia**, Campinas, v. 64, n. 4, p. 577-582, 2005.

ALBUQUERQUE, M. C. C. Quatro séculos de história econômica brasileira. São Paulo, McGraw-Hill do Brasil, 1977.

ALCÁCER, L. Textos de apoio a Química-Física. Determinação da Estrutura Molecular. Métodos Espectroscópios. AEIST, (2007).

ALMEIDA, D. C. B. *et al.* A HERANÇA COLONIAL BRASILEIRA: Quanto as relações sociais e de produção no ciclo do café (1727-2017). **Revista Caribeña de Ciencias Sociales**, 2018.

ALVARADO, R. A.; LINNEMANN, A. R. *The predictive value of a small consumer panel for coffee-cupper Judgment*. **British Food Journal**, vol. 112 no. 9, p. 1023-1032, 2010.

ALVES, J. D. Morfologia do cafeeiro. In: CARVALHO, C. H. S. Cultivares de café. 1 ed. 247 p. Brasília: EMBRAPA, 2007.

ANDRADE, P. B.; LEITÃO, R.; SEABRA, R. M.; OLIVEIRA, M. B.; FERREIRA, M. A. *Coffea canephora variation robusta and Coffea arabica*. **Food Chemistry**, v. 61, n. 511, 1998.

ANDRZEJEWSKI, D.; ROACH, J. A. G.; GAY, M. L.; MUSSER, S. M. *Analysis of Coffee for the Presence of Acrylamide by LC-MS/MS*. **Journal of Agricultural and Food Chemistry**. 52 (7), 1996-2002. 2004.

ANTHONY, F.; BERTRAND, B.; QUIROS, O.; WILCHES, A.; LASHERMES, P.; BERTHAUD, J.; CHARRIER, A. *Genetic diversity of wild coffee (Coffea arabica L.) using molecular markers*. **Euphytica**, v.118, p.53-65, 2001.

ARNAUD, M. J. Em Encyclopedia of Human Nutrition; CABALLERO, B.; SADLER, M. J.; STARIN, J. J., eds.; Academic Press: London, England, 1999, v. 1, p. 206.

ARRUDA, A. C.; MINIM, V. P. R.; FERREIRA, M. A. M.; MINIM, L. A.; SILVA, N. M.; SOARES, C. F. *Justificativas e motivações do consumo e não consumo de café*. **Food Science & Technology**, v. 29, n. 4, p. 754-763, 2009.

ARYA, M.; RAO, L. J. M. *An impression of coffee carbohydrates*. **Critical Reviews in Food Science and Nutrition**. 47 (1), 51-67, 2007.

ASSIS, C.; OLIVEIRA, L. S.; SENA, M. M. *Variable Selection Applied to the Development of a Robust Method for the Quantification of Coffee Blends Using Mid Infrared Spectroscopy*. **Food Analytical Methods**, v. 11, p. 578-588, 2017.

AVALLONE, S.; GUIRAUD, J. P.; GUYOT, B.; OLGUIN, E.; BRILLOUET, J.M. *Polysaccharide constituents of coffee bean mucilage*. **Journal of Food Science**, v. 65, p. 1308-1311, 2000.

BAQUETA, M. R.; ALVES, E. A.; VALDERRAMA, P.; PALLONE, J. A. L. *Brazilian Canephora coffee evaluation using NIRS spectroscopy and discriminant chemometric techniques*. **Journal of Food Composition and Analysis**, v. 116, 2023.

BARBIN, D. F.; FELICIO, A. L. S. M.; SUN, D.; NIXDORF, S. L.; HIROOKA, E. Y. *Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview*. **Food Research International**. v. 61, p. 23-32, 2014.

BARBOSA, L. C. A. Espectroscopia no Infravermelho na caracterização de compostos orgânicos. Viçosa: Editora UFV, 2008. 189 p.

BARTHOLO, G. F.; GUIMARÃES, P. T. G. Cuidados na colheita e preparo do café. **Informe Agropecuário**, Belo Horizonte, v. 18, n. 187, p. 33-42, 1997.

BARTON, F.E. *Analytical application to fibrous foods and commodities*. In: *Near-infrared technology in the agricultural and food industries*. WILLIAMS, P., NORRIS, K., Eds.; **American Association of Cereal Chemists**, Inc.: St. Paul, MN. 1987

BELCHIOR, V.; BOTELHO, B. G.; FRANCA, A. S. *Comparison of Spectroscopy-Based Methods and Chemometrics to Confirm Classification of Specialty Coffees*. **Foods**, v. 11, p. 1655, 2022.

BICCHI, C. P.; PANERO, O. M.; PELLEGRINO, G. M.; VANNI, A. C. *Characterization of roasted coffee and coffee beverages by solid phase microextraction-gas chromatography and principal component analysis*. **Journal Agricultural and Food Chemistry**, v. 45, n. 4680, 1997.

BICHO, N. C. C. *et al.* O Café: Origens, Produção, Processamento e Definição de Qualidade. 1 ed. Lisboa: Escolar Editora, 2011

BONA, E.; MARÇO, P. H.; VALDERRAMA, P. *Chemometrics Applied to Food Control*. **Food Control and Biosecurity**, p. 105–133, 1 jan. 2018.

BORÉM, F. M. Pós-colheita do café. Lavras: Editora UFLA, 2008. 631 p.

BOSSO, H.; BARBALHO, S. M. *Green coffee: economic relevance and a systematic review of the effects on human health*. **Critical Reviews in Food Science and Nutrition**, v. 8, p. 1-17, 2021.

BRAINER, M. S. C. P.; XIMENES, L. F. Produção e mercado do café. **Caderno Setorial ETENE**, ano 6, n. 207, 2021

BRASIL, Ministério da Agricultura, Pecuária e Abastecimento. Portaria SDA nº 570, de 9 de maio de 2022. **Diário Oficial da União**. Publicado em: 11/05/2022, ed. 88, seção 1, p.16.

BRASIL, Ministério da Agricultura, Pecuária e Abastecimento. Regulamento técnico de identificação e de qualidade para a classificação do café beneficiado e de café verde. Instrução normativa nº 8, de 11 de junho de 2003. Gabinete do ministro, 2003.

BRASIL. Ministério da Economia. Secretaria de Comércio Exterior. Balança comercial preliminar parcial do mês – Setembro 2022. 2022. Disponível em: <<https://www.gov.br/produtividade-e-comercio-exterior/pt-br/assuntos/comercio-exterior>>. Acesso em: 12 set. 2022.

BRERETON, R. G. *Chemometrics: Data Driven Extraction for Science*. 2. ed. Hoboken, NJ, USA: John Wiley & Sons, 2018.

BRIANDET, R.; KEMSLEY, K. E.; WILSON, R. H. *Discrimination of Arabica and Robusta in instant coffee by Fourier transform infrared spectroscopy and chemometrics*. **Journal of Agricultural and Food Chemistry**, v. 44, n. 1, p. 170-174, 1996.

BSCA, *Brazil Speciality Coffee Association*. Disponível em: <<https://brazilcoffeenation.com.br/index/home>>. Acesso em: 24 abr. 2021.

BUCKERIDGE, M. S.; TINÉ, M. A. S.; SANTOS, H. P.; LIMA, D. U. Polissacarídeos de reserva da parede celular em semente, estruturas, metabolismo, funções e aspectos ecológicos. **Revista Brasileira de Fisiologia Vegetal**, São Paulo, v. 12, edição especial, p. 137-162, 2000.

BUFFO, R.; CARDELLI-FREIRE, C. *Coffee flavour: a review*. **Flavour and Fragrance Journal**, v. 19, Issue 2, p. 99-104. 2004.

CARACOSTEA, L. M.; SÎRBU, R.; BUȘURICU, F. (2021). *Determination of Caffeine Content in Arabica and Robusta Green Coffee of Indian Origin*. **European Journal of Natural Sciences and Medicine**, 4(1), 69–79.

CARRERA, F.; LEÓN-CAMACHO, M.; PABLOS, F.; GONZÁLEZ A. G. *Authentication of green coffee varieties according to their sterolic profile*. **Analytica Chimica Acta**, v. 370, n. 131, 1998.

CASTILLEJOS-MIJANGOS, L. A.; ACOSTA-CAUDILLO, A.; GALLARDO-VELÁZQUEZ, T.; OSORIO-REVILLA, G.; JIMÉNEZ-MARTÍNEZ, C. *Uses of FT-MIR Spectroscopy and Multivariate Analysis in Quality Control of Coffee, Cocoa, and Commercially Important Spices*. **Foods**. 11, 579. 2022.

CAZARIM, M. de S.; UETA, J. Café: uma bebida rica em substâncias com efeitos clínicos importantes, em especial a cafeína. **Revista de Ciências Farmacêuticas Básica e Aplicada**, v.35, n.3, p.363-370, 2014.

CECAFÉ, Conselho dos Exportadores de Café do Brasil. Disponível em: <<https://www.cecafe.com.br/>>. Acesso em: 24 fev. 2022.

CECAFÉ, Conselho dos Exportadores de Café do Brasil. Relatório mensal – junho 2021. Disponível em: <<https://www.cecafe.com.br/publicacoes/relatorio-de-exportacoes/>>. Acesso em: 03 ago. 2021.

CHARLTON, A. J.; Farrington, W. H. H.; BRERETON, P. *Application of ¹H NMR and multivariate statistics for screening complex mixtures: quality control and authenticity of instant coffee*. **Journal Agricultural and Food Chemistry**, v. 50, n. 3098, 2002.

CHEAH, W. L.; FANG, M. *HPLC-Based Chemometric Analysis for Coffee Adulteration*. **Foods**. 9 (7), 880, 2020.

ČÍŽKOVÁ, H. et al. *Differentiation of coffee varieties according to their sterolic profile*. **Journal of Food and Nutrition Research**, v. 46, n. 1, p. 28–34, 2007.

CLARKE, R. J.; VITZHUM, O. G. **Coffee recent developments**. Oxford: Blackell Science, 2001. 246 p.

CLIFFORD, M. N. *Chlorogenic acids and other cinnamates nature, occurrence and dietary burden*. **Journal of the Science Food and Agriculture**. Oxford, v. 79, p. 363-372, 1999.

CNC, Conselho Nacional do Café. Disponível em: <<https://cncafe.com.br/>>. Acesso em: 07 jan. 2022.

COMBES, M. C.; JOËT, T.; LASHERMES, P. *Development of a rapid and efficient DNA-based method to detect and quantify adulterations in coffee (Arabica versus Robusta)*. **Food Control**, v. 88, p. 198-206, 2018.

CONAB, Companhia Nacional de Abastecimento. Acompanhamento da safra brasileira – safra 2021 – 4º levantamento. v. 8, n. 4, 2021.

Consórcio Pesquisa Café. Sumário Executivo Café, novembro-2021. 2021. Disponível em: <<http://www.consorciopesquisacafe.com.br/index.php/imprensa/noticias/1094-2021-12-27-18-34-41>>. Acesso em: 02 dez. 2021.

CONSTANTINO, L. V.; ZEFFA, D. M.; KOLTUN, A.; URBANO, M. R.; SANTOS SANZOVO; A. W.; NIXDORF, S. L. (2020). *Extraction of soluble sugars from green coffee beans using hot water and quantification by a chromatographic method without an organic solvent*. **Acta Chromatographica**, 32(4), 242-246.

CORRÊA, P. C., OLIVEIRA, G. H. H., BOTELHO, F. M., TRETTO, P. C., ALVES, E. A. (2015) Propriedades físicas e químicas interferentes na pós-colheita do café. In: MARCOLAN, A. L., ESPINDULA, M. C. *Café na Amazônia*. Brasília, DF: Embrapa, p. 401-424.

CORREIA *et al.* *Chemical profiles of Robusta and Arabica coffee by ESI(-)FT-ICR MS and ATR-FTIR: a quantitative approach*. **Analytical Methods**, Issue 42, 8, 7678-7688, 2016.

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

CORREIA, M. N. G. *Influência da torra na evolução dos ácidos clorogênicos do café*. 1990. Lisboa: Dissertação de Doutorado, Instituto Superior de Agronomia, Universidade Técnica de Lisboa.

CORREIA, REDIGYA MEYRELLES. *Perfil químico do Café Capixaba Conilon & Arábica por ESI(-)FT-ICR MS e ATR-FTIR: uma abordagem quantitativa*. 2016. Dissertação (Mestrado em Química) – Universidade Federal do Espírito Santo, Vitória.

COUTO, C. C.; SILVA, O. F.; OLIVEIRA, E. M. M.; SOUSA, C.; CASAL, S. *Near-infrared spectroscopy applied to the detection of multiple adulterants in roasted and ground arabica coffee*. **Foods**, v. 11, n. 1:61, 2021.

COUTO, C. C.; SILVA, O. F.; OLIVEIRA, E. M. M.; SOUSA, C.; CASAL, S. Near-Infrared Spectroscopy Applied to the Detection of Multiple Adulterants in Roasted and Ground Arabica Coffee. *Foods*, 11, 61, 2022.

CRAIG, A. P., FRANCA, A. S., OLIVEIRA, L. S. *Discrimination between Immature and Mature Green Coffees by Attenuated Total Reflectance and Diffuse 115 Reflectance Fourier Transform Infrared Spectroscopy*. *Journal of Food Science*, vol. 76, p. 1162–1168, 2011.

DAVIS, A. P.; CHADBURN, H.; MOAT, J.; O'SULLIVAN, R.; HARGREAVES, S.; LUGHADHA, N. E. *High extinction risk for wild coffee species and implications for coffee sector sustainability*. *Science Advances*, v. 5, n. 1, 2019.

DAVIS, A. P.; RAKOTONASOLO, F. *Six new species of coffee (Coffea) from northern Madagascar*. *Kew Bull*, v. 76, p. 497–511, 2021.

DAVIS, A. P.; TOSH, J.; RUCH, N.; FAY, M. F. *Growing coffee: Psilanthus (Rubiaceae) subsumed on the basis of molecular and morphological data, implications for the size, morphology, distribution and evolutionary history of Coffea*. *Botanical Journal of the Linnean Society*, v. 167, p. 1-21, Dec. 2011.

DEWICK, P. M. *Medicinal Natural Products: A Biosynthetic Approach*. 2 ed. Editora: WILEY, 2001.

DIAS, R. C. E.; BENASSI, M. D. T. *Discrimination between Arabica and Robusta Coffees Using Hydrosoluble Compounds: Is the Efficiency of the Parameters Dependent on the Roast Degree?*. *Beverages*, 2015, 1, 127-139.

DOMINGUES, D. S., PAULI, E. D., ABREU, J. E. M., MASSURA, F. W., CRISTIANO, V., SANTOS, M. J., NIXDORF, S. L. *Detection of roasted and ground coffee adulteration by HPLC by amperometric and by post-column derivatization UV-Vis detection*. *Food Chemistry*, v. 146, p. 353–362, 2014.

DOMINGUES, R.; PEREIRA, S. P.; BACIC, M. J. Hábitos e preferências dos consumidores de café no Brasil, em 2021, comparados com 2019. Instituto Axxus. ABIC, 2021. Disponível em: <https://estatisticas.abic.com.br/estatisticas/pesquisas/habitos-e-preferencias-dos-consumidores-de-cafe/>. Acesso em 23 jan. 2022.

DOWNEY, G.; BOUSSION, J.; BEAUCHÊNE, D. *Authentication of Whole and Ground Coffee Beans by near Infrared Reflectance Spectroscopy*. *Journal Near Infrared Spectroscopy*, 2, 85-92, 1994.

DOWNEY, G.; BRIANDET, R.; WILSON, R. H.; KEMSLEY, E. K. *Near- and Mid-Infrared Spectroscopies in Food Authentication: Coffee Varietal Identification*. *Journal of Agricultural and Food Chemistry*, 45(11), 4357–4361. 1997.

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária. Características das principais variedades de café cultivadas em Rondônia. Documentos/Embrapa Rondônia, ISSN 0103-9865; 93. 21 p., 2004.

ESBENSEN, K. *Multivariate Data Analysis - In Practice*, 5th Edition, CAMO Process AS, Oslo, 2002.

ESTEBAN-DÍEZ, I.; GONZÁLEZ- SÁIZ, J. M.; PIZZARRO, C. *An evaluation of orthogonal signal correction methods for the characterization of Arabica and Robusta coffee varieties. Analytica Chimica Acta*, v. 514, n. 57, 2004.

ESTEBAN-DÍEZ, J. M.; GONZÁLEZ-SÁIZ, C.; SÁENZ-GONZÁLEZ, C; PIZARRO, C. *Coffee varietal differentiation based on near infrared spectroscopy, Talanta*, 71, 221, 2007.

FAO, *Food Outlook: Biannual Report On Global Food Markets*. Rome. 2021. Disponível em: <<https://www.fao.org/documents/card/en/c/cb4479en>>. Acesso em: 13/03/2023.

FARAH, A., SANTOS, T. F. (2015). *The Coffee Plant and Beans: An Introduction*. In: Preedy, V. R. (Ed.). *Coffee in Health and Disease Prevention* (1. ed.). London: Elsevier.

FAZUOLI, L. C. *et al.* Cultivares de café arábica (*Coffea arabica* L.). In: CARVALHO, C. H. S. Cultivares de café. 1 ed. 247 p. Brasília: EMBRAPA, 2007.

FEIRA-MORALES, A. M. *Examining the case of green coffee to illustrate the limitations of grading systems/expert tasters in sensory evaluation for quality control. Food Quality and Preference*, vol. 13 p. 355–367, 2002.

FERNANDES, S. M.; PEREIRA, R. G. F. A.; PINTO, N. A. V. D.; NERY, M. C.; PÁDUA, F. R. M. Constituintes químicos e teor de extrato aquoso de cafés arábica (*Coffea arabica* L.) e conilon (*Coffea canephora* Pierre) torrados. *Ciência e Agrotecnologia*, v. 27, p. 1076–1081, 2003.

FERRÃO, R. G.; FONSECA, A. F. A.; FERRÃO, M. A.; MUNER, L. H. Café Conilon. INCAPER: Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural. 2 ed. 784 p., Espírito Santo: INCAPER, 2017. ISBN 978-85-89274-26-5.

FERREIRA, M. M. C. *Quimiometria: Conceitos, métodos e Aplicações*. Campinas, SP: Editora da Unicamp, 2015.

FERREIRA, T.; GALLUZZI, L.; PAULIS, T.; FARAH, A. *Three centuries on the science of coffee authenticity control. Food Research International*, 149, 110690, 2021.

FISCHER, M.; REIMANN, S.; TROVATO, V.; REDGWELL, R. J. *Polysaccharides of green Arabica and Robusta coffee beans*. **Carbohydrates Research**, v. 330, p. 93-101, 2001.

FONSECA, A. F. A. da; FERRÃO, R. G.; FERRÃO, M. A. G.; VERDIN FILHO, A. C.; VOLPI, P. S. Qualidade do café conilon: operações de colheita e pós-colheita. In: FERRÃO, R. G.; DA FONSECA, A. F. A.; BRAGANÇA, S. M.; FERRÃO, M. A. G.; DE MUNER, L. H. (Eds). *Café Conilon*. Vitória: INCAPER, p. 501 – 520, 2007.

FONSECA, A. F. A. *et al.* Cultivares de café robusta (*Coffea canephora* Pierre ex A. Froehner). In: CARVALHO, C. H. S. *Cultivares de café*. 1 ed. 247 p. Brasília: EMBRAPA, 2007.

FRANK, O.; ZEHENTBAUER, G.; HOFMANN, T. *Bioresponse-guided decomposition of roast coffee beverage and identification of key bitter taste compounds*. **European Food Research Technology**, Heidelberg, v. 222, n. 5/6, p. 492–508, 2006.

GARCIA, L. M. Z.; PAULI, E. D.; VALDERI, C.; CAMARA, C. A. P.; SCARMÍNIO, I. S.; NIXDORF, S. L. *Chemometric evaluation of adulteration profile in coffee due to corn and husk by determining carbohydrates using HPAEC-PAD*. **Journal of Chromatographic Science**, v. 47, p. 825-832, 2009.

GIACALONE, D.; DEGN, T. K.; YANG, N.; LIU, C.; FISK, I.; MÜNCHOW, M. *Common roasting defects in coffee: Aroma composition, sensory characterization and consumer perception*. **Food Quality and Preference**, v. 71, p. 463-474, 2019.

GIRAUDO, A.; GRASSI, S.; SAVORANI, F.; GAVOCI, G.; CASIRAGHI, E.; GEOBALDO, F. *Determination of the geographical origin of green coffee beans using NIRS spectroscopy and multivariate data analysis*. **Food Control**, v. 99, p. 137-145, 2019

GOVERNO DO ESTADO DE SÃO PAULO. Secretaria de Agricultura e Abastecimento - SAA. **Resolução SAA – 28, de 01/06/2007** – Norma Técnica para fixação de identidade e qualidade de café torrado em grão e café torrado e moído.

GRANATO, D.; SANTOS, J. S.; ESCHER, G. B.; FERREIRA, B. L.; MAGGIO, R. M. *Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective*. **Trends in Food Science & Technology**, v. 72, p. 83-90, 2017.

HERRERA, J. C.; LAMBOT, C. *The coffee tree-Genetic diversity and origin*. In: FOLMER, B. (org.). *The Craft and Science of Coffee*, 1 ed., Academic Press, 2016.

HIGSON, S. P. J. *Química Analítica*. São Paulo: McGraw-Hill, 2009.

HOF, M. *Basics of Optical Spectroscopy*. In: GAUGLITZ G.; VO-DINH T. (Ed.) *Handbook of Spectroscopy*. Weinheim: WILEY-VCH, 2003. p. 39-47.

HOTELLING, H. *Analysis of a complex of statistical variables into principal components*, **J. Educ. Psych.**, 24, 417-441, 498-520, (1933).

IBC, Instituto Brasileiro do Café. *Cultura do café no Brasil – Manual de Recomendação*, 4ª edição, 503p, 1981.

IBRAHIM, S. A.; SALAMEH, M. M.; PHETSOMPHOU, S.; YANG, H.; SEO, C. W. *Application of caffeine, 1,3,7-trimethylxantine to control Escherichia coli O157:H7*. **Food Chemistry**, v. 99, n. 4, p. 645-650, 2006.

ICO, *International Coffee Organization*. *Coffee production by exporting countries*. 2021. Disponível em: < https://www.ico.org/trade_statistics.asp?section=Statistics>. Acesso em: 10 jan. 2022.

ICO, *International Coffee Organization*. Disponível em: <<https://www.ico.org/>>. Acesso em: 10 jan. 2022.

ICO, *International Coffee Organization*. *Field practices up characteristics selective US mechanical harvest and wet US dry process: technical unit*. London: Sponsored by the Promotional Fund, 1991. p. 28. (Quality Series Report, 5).

ILLY, A.; VIANI, R. *Espresso coffee: the science of quality*. 2nd ed. San Diego: **Elsevier Academic**, 2005. 398 p.

ISO 11292: 1995. *Instant coffee: Determination of free and total carbohydrate contents. Method using high performance anion-exchange chromatography*. **International Organization for Standardization**. 1995.

JACKSON, J. E. *A Users Guide to Principal Components*, Wiley & Sons Inc., New York, 1991.

IWAI, K.; KISHIMOTO, N.; KAKINO, Y.; MOCHIDA, K.; FUJITA, T. *In vitro antioxidant effects and tyrosinase inhibitory activities of seven hydroxyl cinnamoyl derivatives in green coffee beans*. **Journal of Agricultural and Food Chemistry**. 52, n. 15, p. 4893-4898, 2004.

JESZKA-SKOWRON, M.; SENTKOWSKA, A.; PYRZYŃSKA, K. *et al*. *Chlorogenic acids, caffeine content and antioxidant properties of green coffee extracts: influence of green coffee bean preparation*. **Eur Food Res Technol** 242, 1403–1409 (2016).

KEMSLEY, E. K.; RUAULT, S.; WILSON, R. H. (1995). *Discrimination between Coffea arabica and Coffea canephora variant robusta beans using infrared spectroscopy*. **Food Chemistry**, 54(3), 321–326.

KORNMAN, C. *The Coffee Roaster's Complete Guide to Coffee Varieties and Cultivars*. *Daily Coffee News*. 2019. Disponível em: <https://dailycoffeenews.com/2019/02/07/the-coffee-roasters-complete-guide-to-coffee-varieties-and-cultivars/>. Acesso em: 13 set. 2021.

KOWALSKI, B. R. (1975). *Chemometrics: Views and Propositions*. *Journal of Chemical Information and Modeling*, 15(4), 201–203.

KRAMER, R. *Chemometric techniques for quantitative analysis*. New York, USA. 1998, 203 p.

KRUG, C. A. e CARVALHO, A. Melhoria do cafeeiro: V - Melhoria por hibridação. **Bragantia**. v. 12, n. 4-6, 1952.

LAMBOT, C.; HERRERA, J. C.; BERTRAND, B.; SADEGHIAN, S.; BENAVIDES, P.; GAITA, A. *Cultivating coffee quality-terroir and agro-ecosystem*. In: FOLMER, B. (org.). **The Craft and Science of Coffee**, 1th edition, **Academic Press**, 2016.

LEE, K. G.; SHIBAMOTO, T. *Analysis of volatile components isolated from Hawaiian coffee beans (Coffea arabica L.)*. **Flavour and Fragrance Journal**. v. 17, Issue 5, p. 349-351. 2002.

LEHNINGER, T. M., NELSON, D. L. & COX, M. M. *Princípios de Bioquímica*. 6^a Edição, 2014. Ed. Artmed.

LEROY, T. *et al. Genetics of coffee quality*. *Brazilian Journal of Plant Physiology*. 18 (1): 229-242, 2006.

LIMA FILHO, T.; DELLA LUCIA, S. M.; SARAIVA, S. H.; CARNEIRO, J. C. S.; ROBERTO, C. D. Perfil sensorial e aceitabilidade de bebidas de café tipo expresso preparadas a partir de *blends* de café arábica e conilon. *Enciclopédia Biosfera*, v.7, n.12, 2011.

LIMA FILHO, T.; DELLA LUCIA, S. M.; SARAIVA, S. H.; LEITE, S. T. Qualidade sensorial e físico-química dos cafés arábica e conilon. **Enciclopédia Biosfera**, v.9, n.16, p. 187-1901, 2013.

LINGLE, T. R. *The basics of cupping coffee*. Long Beach: *Specialty Coffee Association of America*, 1993.

LUDWING, I. A., PEÑA, M. P., CID, C., CROZIER, A. *Catabolism of coffee chlorogenic acids by human colonic microbiota*. **BioFactors**. Online, 2013.

MADEIRA, T. B.; SILVA, G. L.; WATANABE, L. S.; LOPES, C. T.; MAZALLI, A. V. G.; ALMEIDA, M. B.; NIXDORF, S. L. Método rápido e sensível para determinação de trigonelina e cafeína por UHPLC-DAD aplicado a café torrado e moído. **SCIENTIA CHROMATOGRAPHICA**, v. 9, p. 160-170, 2017.

MAESO, N.; DEL CASTILLO, C.; CORNEJO, L.; GARCIA-ACICOLLAR, M.; ALGUACIL, L. F.; BARBAS, C. J. *Capillary electrophoresis for caffeine and pyroglutamate determination in coffees study of the in vivo effect on learning and locomotor activity in mice*. **Journal of Pharmaceutical and Biomedical Analysis**. v. 41, n. 4, p. 1095 -1100, 2006.

MAEZTU, L.; SANZ, C.; ANDUEZA, S.; DE PEÑA, M. P.; BELLO, J.; CID, C. *Characterization of espresso coffee aroma by HS-GC-MS and sensory flavor profile*. **Journal Agricultural and Food Chemistry**, v. 49, n. 5437, 2001.

MALAQUIAS, J. V.; CELESTINO, S. M. C.; XAVIER, M. F. F. *Optimization of the roasting conditions of arabica coffee cultivated in the cerrado area of Brazil*. **Brazilian Journal of Food Technology**, Campinas, v. 21, e2016162, 2018.

MALYAR, Y.; MIKHAILENKO, M.; PANKRUSHINA, N.; MIKHEEV, A.; KUZNETSOVA, S.; SHAKHTSHNEIDER, T. (2018). *Effect of Microwave Irradiation on Arabinogalactan and Its Interaction with Betulin Diacetate*. **Russian Journal of Bioorganic Chemistry**.

MAPA, Ministério da Agricultura, Pecuária e Abastecimento. Sumário Executivo Café, junho-2021. 2021. Disponível em: <http://www.sapc.embrapa.br/arquivos/consorcio/informe_estatistico/Sumario_Cafe_junho_2021.pdf>. Acesso em: 02 dez. 2021.

MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. *Multivariate Analysis*, Academic Press Inc, London, 1979.

MARCUCCI, C. T., BENASSI, M. T., ALMEIDA, M. B., NIXDORF, S. L. Teores De Trigonelina, Ácido 5-Cafeoilquínico, Cafeína E Melanoidinas Em Cafés Solúveis Comerciais Brasileiros. **Química Nova**, v. 36 (4), p. 544-548, 2013.

MARQUETTI, I.; Link, J. V.; Lemes, A. L. G.; Scholz, M. B. S.; Valderrama, P.; Bona, E. *Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee*. **Comput. Electron. Agric.** 2016, 121, 313–319

MARTINS, A. L. História do Café. 2 ed. São Paulo: Contexto, 2017.

MATIELLO, J. B. O café: do cultivo ao consumo. São Paulo: Globo, 1991. (Coleção do Agricultor. Grãos).

MAZZAFERA, P. *Trigonelline in coffee*. **Phytochemistry**, v. 30, n. 7, p. 2309-2310, 1991.

MELO, W. L. B. A importância da informação sobre do grau de torra do café e sua influência nas características organolépticas da bebida. São Carlos: Empresa Brasileira de Pesquisa Agropecuária, 2004. 4 p. (Comunicado Técnico, 58).

MENDES, A. N. G. *et al.* Cultivares de importância histórica ou para a pesquisa. In: CARVALHO, C. H. S. Cultivares de café. 1 ed. 247 p. Brasília: EMBRAPA, 2007.

MOGHIMI, A.; MOHAMMAD, H. A.; SAZGARNIA, A.; SARMAD, M. *Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit.* **Biosystems Engineering**, 106 (3), 295–302. 2010.

MONTEIRO, M. C.; TRUGO, L. C. Determinação de compostos bioativos em amostras comerciais de café torrado. **Química Nova**, v. 28, n. 4, p. 637-641, 2005.

MORAES, S. A. L. *et al.* Análise de compostos bioativos, grupos ácidos e da atividade antioxidante do café arábica (*Coffea arabica*) do cerrado e de seus grãos defeituosos (PVA) submetidos a diferentes torras. **Ciência e Tecnologia de Alimentos**, v. 28, p. 198-207, 2008.

MORAIS, S. *et al.* Compostos bioativos e atividade antioxidante do café conilon submetido a diferentes graus de torra. *Química Nova*, São Paulo, v. 32, n. 2, p. 327-331, 2009.

MOREIRA *et al.* *Galactomannans in Coffee* (ch. 19). In: *Coffee in Health and Disease Prevention*. Ed. VICTOR R. PREEDY. Academic Press, p. 173-182. 2015.

MORGANO, M. A.; FARIA, C. G.; FERREIRA, M. M. C. Determinação de açúcar total em café cru por espectroscopia no infravermelho próximo e regressão por mínimos quadrados parciais. **Química Nova**. v. 30, n. 2, p. 346-350, 2007.

MUNYENDO, L.; NJOROGE, D.; HITZMANN, B. *The Potential of Spectroscopic Techniques in Coffee Analysis: A Review.* **Processes**, 10, 71, 2022.

MUSSATTO, S. I., CARNEIRO, L. M., SILVA, J. P.A., ROBERTO, I. C., TEIXEIRA, J. A. *A study on chemical constituents and sugars extraction from spent coffee grounds.* **Carbohydrate Polymers**. v. 83. p. 368–374, 2011

NATELLA, F.; NARDINI, M.; GIANNETTI, I.; DATTILO, C.; SCACCINI, C. *Coffee drinking influences plasma antioxidant capacity in humans.* **Journal of Agricultural and Food Chemistry**, v.50, n. 21, p. 6211-6216, 2002.

NETO, B. D. B.; SCARMINIO, I. S.; BRUNS, R. E. 25 Anos de quimiometria no Brasil. **Química Nova**, v. 29, n. 6, p. 1401–1406, 2006.

NOGUEIRA, M.; TRUGO, L. C. Distribuição de isômeros de ácidos clorogênicos e teores de cafeína e trigonelina em cafés solúveis brasileiros. **Ciência e Tecnologia de Alimentos**. Campinas 23 (2): 296-299, maio-agosto. 2003.

NUNES, F. M.; COIMBRA, M. A. *Influence of polysaccharide composition in foam stability of espresso coffee.* **Carbohydrate Polymers**, 37, p. 283-285, 1998.

NUNES, M. F.; REIS, A.; DOMINGUES, M. R.; COIMBRA, A. *Characterization of galactomannan derivatives in roasted coffee beverages. Journal of Agricultural and Food Chemistry*, Columbus, v. 54, n. 9, p. 3428-3439, 2006.

NYSE, *The New York Stock Exchange*. Disponível em: <<https://www.nyse.com/index>>. Acesso em: 02 dez. 2021.

OBEIDAT, S. M.; HAMMOUDEH, A. Y.; ALOMARY, A. A. *Application of FTIR Spectroscopy for Assessment of Green Coffee Beans According to Their Origin. Journal of Applied Spectroscopy*, v. 84, p. 1051–1055, 2018.

OESTREICH-JANZEN, S. *Chemistry of Coffee. Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*, 2013.

OKUBO, N.; KURATA, Y. *Nondestructive Classification Analysis of Green Coffee Beans by Using Near-Infrared Spectroscopy. Foods* 2019, 8, 82.

OLIVEIRA, R. C. S.; OLIVEIRA, L. S.; FRANCA, A. S.; AUGUSTI, R. *Evaluation of the potential of SPME-GC-MS and chemometrics to detect adulteration of ground roasted coffee with roasted barley. Journal of Food Composition and Analysis*, v. 22, n. 3, p. 257-261. 2009.

OLIVERI, P.; MALEGORI, C.; CASALE, M.; TARTACCA, E.; SALVATORI, G. *An innovative multivariate strategy for HSI-NIRS images to automatically detect defects in green coffee. Talanta*, v. 199, p. 270–276, 2019.

OOSTERVELD, A.; HARMSSEN, J. S.; VORAGEN, A. G. J.; SCHOLS, H. A. *Extraction and characterization of polysaccharides from green and roasted Coffea Arabica beans. Carbohydrate Polymers*, Oxford, v. 52, p. 285-296, 2003.

OOSTERVELD, A.; VORAGEN, A. G. J.; SCHOLS, H. A. (2003). *Effect of roasting on the carbohydrate composition of Coffea arabica beans. Carbohydrate Polymers*. 54(2), 183–192.

OSBORNE, B. G.; FEARN, T. *Applications of near infrared spectroscopy in food analysis*. In: *Near infrared spectroscopy in food analysis*; Longman Scientific & Technical: Harlow, UK., 1986.

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

PASQUINI, C. *Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications. Journal Brazilian Chemistry Society*, v. 14, n2, p. 198-212, 2003.

PATRICIO, F. R. A.; BRAGHINI, M. T.; FAZUOLI, L. C. Resistência de plantas de *coffea arabica*, *coffea canephora* e híbridos interespecíficos à cercosporiose. **Bragantia**. v. 69, n. 4, 2010.

PAULI, E. D.; CRISTIANO, V.; NIXDORF, S. L. Método para determinação de carboidratos empregado na triagem de adulterações em café. **Química Nova**, v. 34, n. 4, p. 689-694, 2011.

PAULI, E. D.; CRISTIANO, V.; NIXDORF, S. L. Método para determinação de carboidratos empregado na triagem de adulterações em café. **Quím Nova**. 34(4): 689–94. 2011

PAULI, E. D.; VALDERI, C.; NIXDORF, S. L. Método para determinação de carboidratos empregado na triagem de adulterações em café. **Química Nova**, São Paulo, v. 34, n. 4, p. 689-694, 2011.

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

PAVIA, D. L.; LAMPMAN, G. M.; KRIZ, G. S. *Introduction to Spectroscopy: A Guide for Students of Organic Chemistry*. Third ed. 2001: Thomson Learning.

PENDERGRAST, M. *Uncommon grounds: the history of coffee and how it transformed our world*. Basic Books: New York. 2010. ISBN: 978-0-465-01836-9.

PENDERGRAST, Mark. *Coffee second only to oil? Is coffee really the second largest commodity?* *Tea & Coffee Trade Journal*. 2009. Disponível em: <<https://www.thefreelibrary.com/Coffee+second+only+to+oil%3F+Is+coffee+really+the+second+largest...-a0198849799>>. Acesso em: 27 out. 2021.

PERRONE, D.; FARAH, A.; DONANGELO, C. M.; DE PAULIS, T.; MARTIN, P. R. *Comprehensive analysis of major and minor chlorogenic acids and lactones in economically relevant Brazilian coffee cultivars*. **Food Chemistry**, Oxford, v. 106, n. 2, p. 859-867, 2008.

PIMENTA, C. J.; VILELA, E. R. Qualidade do café (*Coffea arabica* L.) colhido em sete épocas diferentes na região de Lavras-MG. **Ciência e Agrotecnologia**, Lavras, ed. especial, p. 1481-1491, 2002.

PIZARRO, C.; ESTEBAN-DÍEZ, I.; GONZÁLEZ-SÁIZ, J. M. *Mixture resolution according to the percentage of robusta variety in order to detect adulteration in roasted coffee by near infrared spectroscopy*. **Analytica Chimica Acta**, v. 585, p. 266-276, 2007.

POISSON, L. et al. *The chemistry of roasting: decoding flavor formation*. In: FOLMER, B. *The craft and science of coffee*. Amsterdam: **Academic Press**, 2017. chap. 12, p. 273-309.

PORTILLO, O. R.; ARÉVALO, A. C. *Coffee's carbohydrates. A critical review of scientific literature. Revis Bionatura*. 7(3) 11. 2022.

RÁMON, F. P. *El mundo del café. Ministério de Agricultura, Alimentación y Medio Ambiente*: Madrid, 2014. ISBN: 978-84-491-1424-3.

RANNAR, S.; LINDGREN, F.; GELADI, P.; WOLD, S. *A PLS kernel algorithm for data sets with many variables and fewer objects, Part 1: Theory and Algorithm, J. Chemom.*, 8, 111-125 (1994).

REDGWELL, R. J.; FISCHER, M. *Coffee carbohydrates. Brazilian Journal of Plant Physiology*, v. 18, n. 1, p. 165-174, 2006.

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* 2011, 83, 1352–1358.

RINNAN, Å.; BERG, F. V. D.; ENGELSEN, S. B. *Review of the most common pre-processing techniques for near-infrared spectra. 28(10)*, 1201–1222. 2009.

RODRIGUES, N. P.; BRAGAGNOLO, N. *Identification and quantification of bioactive compounds in coffee brews by HPLC–DAD–MSⁿ. Journal of Food Composition and Analysis*, v. 32 (2), p. 105-115, 2013.

ROGERS, W. J.; MICHAUX, S.; BASTIN, M.; BUCHELI, P. *Changes to the content of sugars alcohols, myo-inositol, carboxylic acids and inorganic anions in developing grains from different varieties of robusta (Coffea canephora) and Arabica (C. arabica) coffees. Plant Science*, v. 149, p. 115-123, 1999.

RUFINO, J. L. S. *Programa Nacional de Pesquisa e Desenvolvimento do Café: antecedentes, criação e evolução. Embrapa Café. Brasília: Embrapa Informação Tecnológica*, 348 p., 2006.

SAHACHAIRUNGRUENG, W.; MEECHAN, C.; VEERACHAT, N.; THOMPSON, A.K.; TEERACHAICHAYUT, S. *Assessing the Levels of Robusta and Arabica in Roasted Ground Coffee Using NIRS Hyperspectral Imaging and FTIR Spectroscopy. Foods*, 11, 3122, 2022.

SAKIYAMA, N. S.; PEREIRA, A. A.; ZAMBOLIM, L. *Melhoramento de café arábica. In: BORÉM, A. (Ed.). Melhoramento de espécies cultivadas. Viçosa: UFV*, 1999. p. 189-204.

SALDAÑA, M. D. A.; MAZZAFERA, P.; MOHAMED, R. S. *Extração dos alcaloides: cafeína e trigonelina dos grãos de café com CO₂ supercrítico. Ciência e Tecnologia de Alimentos*, v. 17, n. 4, p. 371-376, 1997.

SALVA, T. J. G.; LIMA, V. B. A composição química do café e as características da bebida e do grão. **O agrônomo**, Campinas, v. 59, n. 1, p 57-59, 2007

SANTOS, J. R.; SARRAGUÇA, M. C.; RANGEL, A. O. S. S.; LOPES, J. A. *Evaluation of green coffee beans quality using near infrared spectroscopy: A quantitative approach*. **Food Chemistry**, v. 135, Issue 3, p. 1828-1835. 2012.

SAVITZKY, A.; GOLAY, M. J. E. *Smoothing and Differentiation of Data by Simplified Least Squares Procedures*. **Analytical Chemistry**. 36(8), 1627-1639. 1964.

SCAA, *Specialty Coffee Association of America*. *SCAA Protocols – Cupping Specialty Coffee*, v. 16, 2015.

SCHENKER, S.; ROTHGEB, T. *The toast- Creating the beans' signature*. In: FOLMER, B. (org.). **The Craft and Science of Coffee**, 1th ed, Academic Press, 2016.

SCHOLZ, M. B. S. Tipologia dos cafés paranaense: uma abordagem através da análise fatorial múltipla dos aspectos físico-químicos e sensoriais. 2008. Tese (Doutorado em Ciência de Alimentos) - Universidade Estadual de Londrina, Londrina.

SCHOLZ, M. B. S.; FIGUEIREDO, V. R. G.; SILVA, J. V. N.; KITZBERGER, C. S. G. Características físico-químicas de grãos verdes e torrados de cultivares de café (*Coffea arabica* L.) do IAPAR. **Coffee Science**, v. 6, 245, 2011.

SCHOLZ, M. B. S.; KITZBERGER, C. S. G.; PEREIRA, L. F. P.; DAVRIEUX, F.; POT, D.; CHARMETANTD, P.; LEROY, T. *Application of near infrared spectroscopy for green coffee biochemical phenotyping*. **J. Near Infrared Spectrosc.** 2014, 22, 411–421.

SECOUARD, S.; GRISEL, M.; MALHIAC, C. *Flavour release study as a way to explain xanthan–galactomannan interactions*. **Food Hydrocolloids**, 21(8), 1237–1244. 2007

SENAR, Serviço Nacional de Aprendizagem Rural. Café: classificação e degustação. Serviço Nacional de Aprendizagem Rural. Brasília: SENAR, 2017, 112 p.

SENINDE, D. R.; CHAMBERS, E. *Coffee Flavor: A Review*. **Beverages**, 6 (3), 44, 2020.

SEVERINI, C., DEROSI, A., RICCI, I., FIORE, A. G., CAPORIZZI, R. *How Much Caffeine in Coffee Cup? Effects of Processing Operations, Extraction Methods and Variables*. In: LATOSINSKA, J. N., LATOSINSKA, M. *The Question of Caffeine*. London: IntechOpen, 2017.

SEZER, B.; APAYDIN, H.; BILGE, G.; BOYACI, I. H. (2018). *Coffee arabica Adulteration: Detection of Wheat, Corn and Chickpea*. **Food Chemistry**, v. 264, p. 142-148, 2018.

SEZER, B.; APAYDIN, H.; BILGE, G.; BOYACI, I. H. *Coffee arabica adulteration: Detection of wheat, corn and chickpea. Food Chemistry*, v. 264, p. 142–148, 2018.

SHI, H.; YU, P. *Comparison of grating-based near-infrared (NIRS) and Fourier transform mid-infrared (ATR-FT/MIR) spectroscopy based on spectral preprocessing and wavelength selection for the determination of crude protein and moisture content in wheat. Food Control*, v. 82, p. 57–65, 2017.

SILVERSTEIN, M. R.; WEBSTER, F. X. *Identificação Espectrométrica de Compostos Orgânicos*. 6 ed. Rio de Janeiro: LTC – Livros Técnicos e Científicos Editora, 2000. 460 p.

SIVETZ, M.; DESROSIER, N. W. *Physical and chemical aspects of coffee. Coffee Technology*, West por, p. 527-575, 1979.

SKOOG, D. A.; HOLLER, F. J.; CROUCH, S. R. *Princípios de Análise Instrumental*. 6a. ed. Porto Alegre: Bookman, 2009. p. 1056

SUN, X.; ZHU, K.; LIU, J. *Nondestructive detection of reducing sugar of potato flours by near infrared spectroscopy and kernel partial least square algorithm. Journal of Food Measurement and Characterization*, 13, 231-237. 2019.

TOCI, A. T.; FARAH, A.; PEZZA, H. R.; PEZZA, L. *Coffee adulteration: more than two decades of research. Critical Reviews in Analytical Chemistry*, v. 46:2, p. 83-92, 2016.

TOLESSA, K.; RADEMAKER, M.; BAETS, B.; BOECKX, P. *Prediction of specialty coffee cup quality based on near infrared spectra of green coffee beans. Talanta*, v. 150, 2016, p. 367-374. 2016.

TOPIK, S. *A Presença do estado na Economia Política do Brasil. 1889 a 1930*. São Paulo: Record, 1987.

TUGNOLO, A.; GIOVENZANA, V.; MALEGORI, C.; OLIVERI, P.; CASSON, A.; CURATITOLI, M.; GUIDETTI, R.; BEGHI, R. *A reliable tool based on near-infrared spectroscopy for the monitoring of moisture content in roasted and ground coffee: A comparative study with thermogravimetric analysis. Food Control* 2021, 130, 108312.

URYUPIN, A. B.; PEREGUDOV, A. S. *Application of NMR techniques to the determination of the composition of tobacco, coffee, and tea products. Journal of Analytical Chemistry*, v. 68, p. 1021–1032, 2013.

USDA, *United States Department of Agriculture. Coffee: world markets and trade*. Disponível em: <<https://usda.library.cornell.edu/concern/publications/m900nt40f?locale=en>>. Acesso em: 29 dez. 2021.

VENDRAME, I. R.; FILHO, I. L. S. Análise dos requisitos da qualidade do café na opinião dos consumidores: Um estudo em uma indústria cafeeira no estado da Bahia, 2019. Monografia (Bacharel em Engenharia de Produção) – Universidade Tecnológica Federal do Paraná Medianeira, 2019.

WANG, X.; LIM, L. *Physicochemical Characteristics of Roasted Coffee*. In: PREEDY, Victor R. **Academic Press** (ed.). **Coffee in Health and Disease Prevention**, p. 247-254, 2015.

WANG, Y.; JIN, G.; LI, L.; LIU, Y.; KIANPOOR, Y. K.; NING, J.; ZHANG, Z. *NIR hyperspectral imaging coupled with chemometrics for nondestructive assessment of phosphorus and potassium contents in tea leaves*. **Infrared Physics & Technology**, 108. 2020.

WEI, F.; FURIHATA, Z.; KODA, M.; HU, F.; KATO, R.; MIYAKAWA, T.; TANOKURA, M. *¹³C NMR-Based Metabolomics for the Classification of Green Coffee Beans According to Variety and Origin*. **Journal of Agricultural and Food Chemistry**, 60 (40), 10118-10125, 2012.

WOLD, S. *Chemometrics; what do we mean with it, and what do we want from it?* **Chemometrics and Intelligent Laboratory Systems**. 30 (1), 109-115, 1995.

XIAOBO, Z.; JIEWEN, Z.; POVEY, M. J.W.; HOLMES, M.; HANPIN, M. Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta*, v. 667, p. 14–32, 2010, *apud* FERREIRA, D. S. Marcadores espectrais no visível e infravermelho próximo para identificação de sementes em progênies de *Ricinus communid* L. 2015. 134 f. Tese (Doutorado em Agronomia) – Universidade Federal da Paraíba, João Pessoa, Paraíba, 2015.

ZAREEF, M.; CHEN, Q.; HASSAN, M. M.; ARSLAN, M.; HASHIM, M. M.; AHMAD, W.; KUTSANEDZIE, F. Y. H.; AGYEKUM, A. A. *An overview on the applications of typical nonlinear algorithms coupled with NIRS spectroscopy in food analysis*. **Food Engineering Reviews**, v.12, p. 173-190, 2020.

ANEXO A

Tabela 10. Amostras utilizadas no estudo, com dados de espécie, defeitos e tipo, segundo metodologia da COB

Nº	Codificação	Amostra	Estado	Espécie	Defeitos	Tipo
1	A1	BA026	Bahia	Arábica	310	7-35
2	A2	BA027	Bahia	Arábica	130	6-60
3	A3	BA028	Bahia	Arábica	278	7-30
4	A4	BA029	Bahia	Arábica	173	7-05
5	A5	BA030	Bahia	Arábica	70	5-35
6	A6	BA031	Bahia	Arábica	194	7-10
7	A7	BA032	Bahia	Arábica	99	6-10
8	A8	BA033	Bahia	Arábica	188	7-10
9	A9	BA034	Bahia	Arábica	235	7-20
10	A10	BA035	Bahia	Arábica	65	5-25
11	A11	ES004	Espírito Santo	Arábica	28	4
12	A12	MG043	Minas Gerais	Arábica	111	6-15
13	A13	MG049	Minas Gerais	Arábica	38	4-30
14	A14	MG052	Minas Gerais	Arábica	53	5-10
15	A15	MG055	Minas Gerais	Arábica	21	3-30
16	A16	MG060	Minas Gerais	Arábica	86	6
17	A17	MG069	Minas Gerais	Arábica	125	6-30
18	A18	MG073	Minas Gerais	Arábica	205	7-10
19	A19	MG084	Minas Gerais	Arábica	67	5-30
20	A20	MG120	Minas Gerais	Arábica	171	7
21	A21	MG131	Minas Gerais	Arábica	154	6-45
22	A22	PR016	Paraná	Arábica	258	7-25
23	A23	PR017	Paraná	Arábica	161	7
24	A24	PR018	Paraná	Arábica	285	7-30
25	A25	PR021	Paraná	Arábica	208	7-10
26	A26	PR022	Paraná	Arábica	313	7-35
27	A27	PR023	Paraná	Arábica	113	6-15
28	A28	PR024	Paraná	Arábica	159	6-45

29	A29	SP092	São Paulo	Arábica	155	6-45
30	A30	SP094	São Paulo	Arábica	71	5-35
31	A31	SP096	São Paulo	Arábica	176	7-05
32	A32	SP097	São Paulo	Arábica	129	6-30
33	A33	SP099	São Paulo	Arábica	43	4-45
34	A34	SP100	São Paulo	Arábica	146	6-30
35	A35	SP101	São Paulo	Arábica	63	5-20
36	A36	SP102	São Paulo	Arábica	88	6
37	A37	SP114	São Paulo	Arábica	5	2-10
38	A38	SP116	São Paulo	Arábica	8	2-25
39	C1	ES002	Espírito Santo	Canéfora	180	7-05
40	C2	ES003	Espírito Santo	Canéfora	332	7-40
41	C3	ES005	Espírito Santo	Canéfora	129	6-30
42	C4	ES006	Espírito Santo	Canéfora	120	6-25
43	C5	ES105	Espírito Santo	Canéfora	185	7-05
44	C6	ES106	Espírito Santo	Canéfora	286	7-30
45	C7	ES107	Espírito Santo	Canéfora	354	8
46	C8	ES108	Espírito Santo	Canéfora	175	6-40
47	C9	ES110	Espírito Santo	Canéfora	115	6-20
48	C10	ES111	Espírito Santo	Canéfora	186	7-05
49	C11	ES112	Espírito Santo	Canéfora	48	5-05
50	C12	ES142	Espírito Santo	Canéfora	102	6-10
51	C13	ES143	Espírito Santo	Canéfora	194	7-05
52	C14	ES145	Espírito Santo	Canéfora	357	7-45
53	C15	ES146	Espírito Santo	Canéfora	254	7-20
54	C16	ES147	Espírito Santo	Canéfora	22	7-10
55	C17	RO002	Rondônia	Canéfora	111	6-15
56	C18	RO003	Rondônia	Canéfora	57	5-10
57	C19	RO004	Rondônia	Canéfora	152	6-45
58	C20	RO005	Rondônia	Canéfora	158	6-35
59	C21	RO006	Rondônia	Canéfora	24	3-45
60	C22	RO007	Rondônia	Canéfora	349	7-45
61	C23	RO008	Rondônia	Canéfora	137	6-35

62	C24	RO009	Rondônia	Canéfora	190	7-10
63	C25	RO010	Rondônia	Canéfora	146	6-40
64	C26	RO011	Rondônia	Canéfora	162	7
65	C27	RO013	Rondônia	Canéfora	265	7-25
66	C28	RO015	Rondônia	Canéfora	181	7-05
