

UNIVERSIDADE FEDERAL DO PARANÁ

MARCOS ANTONIO TEIXEIRA VOLOCHEN

IDENTIFICAÇÃO DE ADULTERANTES PRESENTES EM CAFÉ TORRADO MOÍDO  
POR EQUIPAMENTO DE IMAGEAMENTO  
HIPERESPECTRAL DE BAIXO CUSTO E FERRAMENTAS QUIMIOMÉTRICAS

CURITIBA

2024

MARCOS ANTONIO TEIXEIRA VOLOCHEN

IDENTIFICAÇÃO DE ADULTERANTES PRESENTES EM CAFÉ TORRADO MOÍDO  
POR EQUIPAMENTO DE IMAGEAMENTO  
HIPERESPECTRAL DE BAIXO CUSTO E FERRAMENTAS QUIMIOMÉTRICAS

Dissertação apresentada ao curso de Pós-Graduação em Química, Setor de Ciência Exatas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Química.

Orientador: Prof. Dr. Prof. Dr. Frederico L. F. Soares

CURITIBA

2024

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)  
UNIVERSIDADE FEDERAL DO PARANÁ  
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIA E TECNOLOGIA

Volochen, Marcos Antonio Teixeira

Identificação de adulterantes presentes em café torrado moído por equipamento de imageamento hiperespectral de baixo custo e ferramentas quimiométricas / Marcos Antonio Teixeira Volochen. – Curitiba, 2024.

1 recurso on-line : PDF.

Dissertação (Mestrado) - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Química.

Orientador: Frederico L. F. Soares

1. Café. 2. Imageamento Hiperespectral. 3. Quimiometria. I. Universidade Federal do Paraná. II. Programa de Pós-Graduação em Química. III. Soares, Frederico L. F.. IV . Título.

Bibliotecário: Leticia Priscila Azevedo de Sousa CRB-9/2029

## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação QUÍMICA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **MARCOS ANTONIO TEIXEIRA VOLOCHEN** intitulada: **IDENTIFICAÇÃO DE ADULTERANTES PRESENTES EM CAFÉ TORRADO MOÍDO POR EQUIPAMENTO DE IMAGEAMENTO HIPERESPECTRAL DE BAIXO CUSTO E FERRAMENTAS QUIMIOMÉTRICAS**, sob orientação do Prof. Dr. FREDERICO LUIS FELIPE SOARES, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 27 de Março de 2024.

Assinatura Eletrônica

01/04/2024 08:42:41.0

FREDERICO LUIS FELIPE SOARES

Presidente da Banca Examinadora

Assinatura Eletrônica

28/03/2024 14:00:34.0

JOSÉ MANUEL AMIGO RUBIO

Avaliador Externo (UNIVERSITY OF THE BASQUE COUNTRY)

Assinatura Eletrônica

28/03/2024 14:00:18.0

BRUNO JOSÉ GONÇALVES DA SILVA

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

## AGRADECIMENTOS

A minha companheira, Emanuele Cristina dos Santos Pohlod, pelo incentivo, amor, amizade e apoio incondicional.

Aos meus pais Antonio Volochen e Margarete Etelvina Teixeira Volochen pela compreensão e suporte.

A minha irmã, Vanessa Volochen, pelo companheirismo, apoio e empolgação.

Ao meu orientador, Professor Dr. Frederico Soares, pela orientação, conselhos paciência, e dedicação.

Ao Guilherme Post Sabin pelo financiamento e total suporte científico para desenvolvimento deste projeto.

Ao Ministério da Agropecuária, Pecuária e Abastecimento, pela parceria e confiança ao conceder as amostras usadas neste projeto.

Aos amigos de laboratório Priscila Lagner, Luigy Tschoke, Beatriz Cestaro, Leonardo Lemes pelo companheirismo e acolhimento.

E aos órgãos de órgãos de fomento.

## RESUMO

Durante o bienio 2019/20, cerca de 165,0 milhões de sacas de 60kg de café verde foram produzidas no Brasil. Sendo o café uma das bebidas quentes mais consumidas do mundo, garantir a qualidade do produto final, evitando adulterações é de extrema relevância. Com isso, o governo brasileiro já atua criando legislações para o controle do café torrado moído consumido no país. Todavia, não há uma metodologia padrão para esse controle, restando ao analista realizar a contagem manual da presença de adulterantes na amostra. Além de ser uma prática demorada, a necessidade de um analista identificar visualmente a diferença entre grãos parecidos traz grande subjetividade, sendo facilmente contextada em laudos técnicos. Neste trabalho, foi desenvolvido uma metodologia de imageamento hiperespectral de baixo custo na região espectral entre 350 a 920nm para analisar amostras de café torrado moído misturadas com os adulterantes casca de café, milho, cevada, caroço de açaí e trigoilho entre concentrações de 3% a 20% (m/m) e sem nenhum tratamento. A área analisada foi de 5x5 mm de cada amostra, com resolução espacial de 0,2mm / pixel. Usando os espectros médios gerados por cada imageamento, foi possível identificar o tipo de café majoritário presente na amostra com mais de 95% de assertividade usando ferramentas de discriminação como PLS-DA e DD-SIMCA. Analisando cada espectro da imagem, a alta resolução vinculada a ferramentas quilométricas não-supervisionadas (MCR-ALS) e supervisionadas de discriminação (PLS-DA, DD-SIMCA) gerou mapas de superfície com clara distinção entre as componentes das misturas, café e adulterante, localizando espacialmente a possível adulteração presente na amostra.

Palavras-chave: Coffee; HSI; MCR-ALS; DD-SIMCA; PLS-DA.

## ABSTRACT

During the biennium 2019/20, approximately 165.0 million 60kg sacks of green coffee were produced in Brazil. As coffee is one of the most consumed hot beverages in the world, ensuring the quality of the final product, avoiding adulteration, is of utmost importance. Consequently, the Brazilian government is already taking action by creating legislation to control the roasted ground coffee consumed in the country. However, there is no standard methodology for this control, leaving the analyst to manually count the presence of adulterants in the sample. Besides being a time-consuming practice, the need for an analyst to visually identify the difference between similar grains introduces significant subjectivity, easily contested in technical reports. In this study, a low-cost hyperspectral imaging methodology was developed in the spectral range between 350 to 920nm to analyze samples of roasted ground coffee mixed with adulterants such as coffee husks, corn, barley, acai seeds, and cracked wheat at concentrations ranging from 3% to 20% (m/m) without any treatment. The analyzed area was 5x5 mm for each sample, with a spatial resolution of 0.2 mm/pixel. Using the average spectra generated by each imaging, it was possible to identify the predominant type of coffee present in the sample with over 95% accuracy using discrimination tools such as PLS-DA and DD-SIMCA. Analyzing each spectrum of the image, the high resolution coupled with unsupervised multivariate curve resolution (MCR-ALS) and supervised discrimination tools (PLS-DA, DD-SIMCA) generated surface maps with clear distinction between the components of the mixtures, coffee and adulterant, spatially locating the possible adulteration present in the sample.

Keywords: Coffee; HSI; MCR-ALS; DD-SIMCA; PLS-DA.

## LISTA DE FIGURAS

Figura 1 - Produção mundial de café 2018. ....	18
Figura 2 –Propriedades produtoras de café do tipo Arábica e Robusta no Brasil. ....	19
Figura 3 –Discos de cor e escala numérica referente ao grau de torra definido pela SCA.....	22
Figura 4 - Redução das dimensões dos dados iniciais. ....	27
Figura 5 - Gráfico de aceitação DD-SIMCA. Círculos verdes são amostras dentro do grupo alvo e quadrados vermelhos são amostras de um grupo alternativo.....	33
Figura 6 - Amostras enviadas pelo MAPA.....	38
Figura 7 - Espectrômetro (círculo amarelo) acoplado a mesa CNC (azul). ....	39
Figura 8 - Configuração do equipamento de imageamento hiperespectral de baixo custo desenvolvido no projeto de mestrado. ....	40
Figura 9 - Ordem de execução do equipamento de imageamento hiperespectral, ...	41
Figura 10 - Espectros cedidos pelo MAPA das amostras enviadas. ....	44
Figura 11 - Gráfico de predição de Y predito para grupo Café puro, em destaque as amostras de cada tipo de metodologia de torrefação. ....	46
Figura 12 - Espectros UV-Vis médio do café arábica e café conilon em linha sólida, e os adulterantes em linha tracejadas. Do lado esquerdo dados brutos e ao lado direito com tratamento normalização SNV. ....	47
Figura 13 - Escores da PCA entre as amostras de café arábica(vermelho) e café conilon(verde). ....	48
Figura 14 - Gráfico de aceitação do modelo DD-SIMCA café arábica para as amostras adulteradas. Pontos em verde indicam classificação da amostra como arábica, pontos em vermelho indicam classificação como não arábica.....	52
Figura 15 - Simulação de aquisição de espectros em 3 posições diferentes da amostra (círculos amarelos). A diferença na colocação dos pixels indicam composições. ....	54
Figura 16 - Loadings dos modelos DD-SIMCA para a)café arábica e b)café conilon. ....	55
Figura 17 - Fluxograma modelos DD-SIMCA. ....	57

Figura 18 - Área de aceite calculada pelo modelo DDS-café arábica. Os pontos verdes indicam espectros semelhança com modelo, pontos vermelhos indicam espectros diferente de café arábica.....	60
Figura 19 - Imageamento da amostra com predição do modelo DD-SIMCA café arábica. ....	60
Figura 20 - Quantidade de amostras indicadas como puras e adulteradas por seus respectivos modelos DD-SIMCA.....	61
Figura 21 - Mapas preditos pelos respectivos modelos de café das amostras. Em azul são áreas atribuídas ao tipo de café, e em vermelho pontos desconhecidos. ....	61
Figura 22 – Corte espectral representado pela área vermelha. ....	63
Figura 23 - Vetor de regressão dos modelos PLS-DA. Em azul para café arábica e em laranja para café conilon. ....	66
Figura 24 - Mapa de predição da amostra de café arábica e milho pelo modelos PLS-DA de café arábica.....	67
Figura 25 - Desempenho da classificação das amostras por seus respectivos modelos PLS-DA.....	67
Figura 26 - Mapas de predições PLS-DA para café arábica. ....	68
Figura 27 - Mapas de predições PLS-DA para café conilon.....	68
Figura 28 - Imagem simulada de uma amostra de café. Em marrom, temos café e os pontos mais escuros, tendendo a preto, são de adulterantes. O círculo amarelo no meio é área de coleta dos espectros.....	70
Figura 29 - Configuração dos MCR-ALS executados e testes de similaridade com o perfil espectral recuperado.....	72
Figura 30 - Comparação entre espectros sem alisamento e com alisamento.....	72
Figura 31 - Perfil espectral(a) e mapa de concentração(b) recuperada pelo MCR-ALS da amostra id_9732422, mistura de café arábica com milho, com pré-tratamento SNV.....	73
Figura 32 - Mapas de concentração recuperadas a partir de diferentes pré-tratamentos da amostra id_9732422.....	75
Figura 33 - Mapa do perfil de concentração recuperado pelo MCR-ALS com suas respectivas componentes. Abaixo de cada mapa de cor está a atribuição de cada componente em função dos testes de similaridade. ....	77

Figura 34 - MCR-ALS de 5 amostras de misturas terciarias. ....	78
Figura 35 - Mapas recuperados por teste não-supervisionados e supervisionados da amostra de café arábica mais milho.....	80
Figura 36 - Mapas recuperados por teste não-supervisionados e supervisionados da amostra de mistura terciaria.....	80

## LISTA DE TABELAS

Tabela 1. PARÂMETROS DE QUALIDADE DO CAFÉ TORRADO .....	20
Tabela 2 – Tabela de confusão a partir de um modelo PLS-DA genérico.....	30
Tabela 3 - Tabela de confusão do modelo PLS-DA para verificar a adulteração de café na região do NIR. ....	45
Tabela 4 - Tabela de confusão da calibração do modelo PLS-DA 1.....	50
Tabela 5 - Resultados modelo DD-SIMCA para café arábica. ....	52
Tabela 6 - Resultados modelo DD-SIMCA para café conilon.....	53
Tabela 7 - Número de espectros usados na construção e validação de cada um dos modelos DD-SIMCA.....	55
Tabela 8 - Resultados referente a amostra de café arábica e milho (id : 9732422) ao ser aplicado nos modelos DD-simca.....	58
Tabela 9 - Configuração do conjunto de dados usados para calibração e validação dos modelos.....	63
Tabela 10 - Figuras de mérito para modelos PLS-DA para café arábica. ....	64
Tabela 11 - Figuras de mérito para modelos PLS-DA para café conilon.....	65
Tabela 12 - Teste de similaridade dos espectros recuperados a partir de cada pré- tratamento. ....	75
Tabela 13 - Figura de mérito com os valores referente aos piores desempenhos para cada pré-tratamento.....	76
Tabela 14 - Eficiência em identificar corretamente as componentes recuperadas por MCR-ALS com teste de similaridade de amostras com misturas binárias. ....	76
Tabela 15 - Eficiência em identificar corretamente as componentes recuperadas por MCR-ALS com teste de similaridade de amostras com misturas ternárias.....	78

## LISTA DE ABREVIATURAS OU SIGLAS

ICA	- International Coffee Organization
MAPA	- Ministério da Agropecuária, Pecuária e Abastecimento
SDA	- Secretária de Defesa Agropecuária
NIR	- Near infrared
HSI	- Hyperspectral image
PCA	- Principal Component Analysis
PLS	- Partial least squares
PLS-DA	- Partial least squares Discriminant Analysis
DD-SIMCA	- Data Driven-Soft Independent Modeling of Class Analogy
MCR-ALS Squares	- Multivariate Curve Resolution – Alternating Least

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>16</b>
<b>2 REVISÃO DE LITERATURA</b> .....	<b>18</b>
2.1 CAFÉ .....	18
2.2 CONTROLE DE QUALIDADE DO GRÃO DE CAFÉ .....	19
2.2.1 Espectroscopia ultravioleta-visível (Uv-VIS) .....	21
2.2.2 Espectroscopia infravermelho próximo (NIR) .....	23
2.2.3 Imagens hiperespectral .....	24
2.3 QUÍMIOMETRIA .....	25
2.3.1 Análise por componentes principais (PCA) .....	26
2.3.2 Mínimos quadrados parciais (PLS) .....	28
2.3.2.1 Mínimos quadrados parciais com análise discriminantes (PLS-DA) .....	30
2.3.3 Data driven-soft independent modeling of class analogy (DD-SIMCA) .....	31
2.3.4 Resolução de curva multivariada por mínimo quadrados alternados (MCR-ALS)	
34	
<b>3 OBJETIVOS GERAIS E ESPECÍFICOS</b> .....	<b>37</b>
3.1 OBJETIVO GERAL .....	37
3.2 OBJETIVO ESPECÍFICO .....	37
<b>4 MATERIAIS E MÉTODOS</b> .....	<b>38</b>
<b>5 RESULTADOS E DISCUSSÃO</b> .....	<b>44</b>
5.1 MODELO COM ESPECTROS NIR CEDIDOS PELO MAPA .....	44
5.2 ANÁLISE EXPLORATÓRIA DOS DADOS DE HSI-VIS .....	47
5.3 MODELOS COM ESPECTROS MÉDIOS .....	49
5.3.1 Modelos PLS-DA com espectros médios .....	49
5.3.2 Modelos DD-SIMCA com espectros médios .....	51
5.4 IMAGEAMENTO HIPERESPECTRAL UV-VIS .....	53
5.4.1 Modelos DD-SIMCA .....	54
5.4.1.1 Aplicando os modelos DD-SIMCA nas amostras adulteradas .....	57
5.4.2 Modelos PLS-DA .....	62
5.4.2.1 Aplicando os modelos PLS-DA nas amostras adulteradas .....	66
5.4.3 MCR-ALS .....	70
5.4.4 Comparação entre os mapas recuperados: PLS-DA vs. DD-SIMCA vs. MCR-ALS	
79	

<b>6 CONCLUSÃO .....</b>	<b>82</b>
<b>REFERÊNCIAS.....</b>	<b>84</b>

## 1 INTRODUÇÃO

O café é uma das bebidas mais populares do mundo. Seu consumo mundial chega aproximadamente a 1,3 xícaras de café por dia. O Brasil tem papel de destaque na produção do grão, sendo o maior exportador do mundo, chegando a uma produção de 165,0 milhões de sacas de 60kg de café verde no biênio de 2019/20.<sup>1 2</sup>

Devido tamanha popularidade da bebida, medidas para assegurar a segurança alimentar do produto, evitando produtos danificados e/ou adulterados são necessários para garantir um experiência de consumo agradável para o usuário.

Assim, o Brasil vem atuando na criação de legislações para controle da qualidade do produto final distribuído em solo nacional. A portaria Nº 570, de 9 de maio de 2022 da Secretária de Defesa Agropecuária (SDA), do Ministério da Agropecuária, Pecuária e Abastecimento (MAPA) publicou em diário oficial as condições quais o café torrado moído deve atender para ser considerado próprio para consumo. Entre elas, o limite de 1% (m/m) de impurezas relacionadas ao processo produtivo do café torrado moído, como paus e cascas provenientes da lavoura, e a completa ausência de elementos estranhos com indicativos de ato ilícito de adulteração. Entretanto, a portaria é vaga ao definir o método de avaliação desses parâmetros, deixando em aberto a metodologia que deve ser seguida para verificar tais condições.<sup>3</sup>

Na literatura, a Fundação Ezequiel Dias apresenta um método para quantificação de adulterantes nas amostras de café torrado e moído com auxílio de microscópios. Nela, a amostra de café torrado moído passa por uma lavagem em solvente orgânico, e após seca, uma alíquota é analisada em um microscópio, onde um analista verifica se existe a presença de adulterantes.<sup>4</sup> Esse método, além de demorado, pode trazer consigo a subjetividade do analista, já que é preciso identificar visualmente o grão de adulterante de coloração extremamente similar ao da amostra de café torrado moído.

Por conta da subjetividade desta análise diversas metodologias analíticas tem sido propostas para se avaliar a presença de adulterantes em amostras de café, sendo as técnicas cromatográficas as mais difundidas. No que tange as técnicas espectroscópicas, alguns trabalhos foram reportados na literatura<sup>5,6</sup>, tentando facilitar o processo e aumentar a frequência analítica da metodologia, porém apesar

de promissores, boa parte dos erros de predição relatados ainda estão acima do que a legislação brasileira impõe, <1% de elementos estranhos.

Com isso, a técnica de imageamento hiperespectral (em inglês, hyperspectral image, sigla, HSI) se torna uma alternativa para encontrar valores de adulterantes dentro dos limites preconizados. Ao mapear toda uma área pré-selecionada da amostra, diminui-se as chances de não se obter espectros referente a uma das componentes das misturas, aumentando assim a sensibilidade da técnica espectroscópica.

Desta forma, a técnica de HSI, com auxílio de ferramentas quimiométricas, entrega uma inspeção visual padronizada e fundamentada estatisticamente, eliminando qualquer subjetividade humana. E ademais, após analisada e identificada a localização do adulterante, caso haja necessidade, pode haver testes complementares para garantir a identificação do adulterante.

Com isso, esse trabalho busca desenvolver uma metodologia utilizando um equipamento de HSI de baixo custo na região espectral de 320 a 920nm afim de identificar a presença de adulterantes como casca de café, milho, cevada, caroço de açaí e trigoilho presentes em amostra de café torrado e moído.

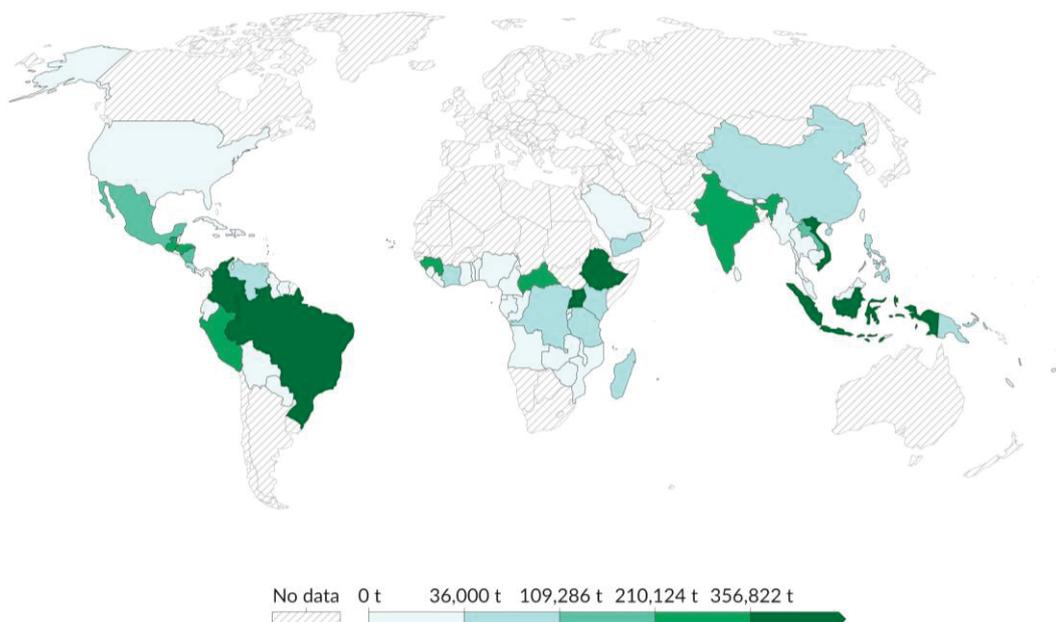
## 2 REVISÃO DE LITERATURA

### 2.1 CAFÉ

O café é uma das bebidas quentes mais populares do mundo, tendo um consumo mundial de cerca de 173,1 milhões de sacos de 60kg de café, de acordo com a Organização Internacional de Café (*International Coffee Organization – ICO*).<sup>7</sup> O mercado mundial do café fresco e instantâneo, movimentou cerca de \$70,86 bilhões em 2011, com 7,9 milhões de toneladas produzidas, onde 6,2 milhões de toneladas (78%) foram dedicadas a exportação.<sup>2</sup> Um valor mais atualizado mostra um crescimento na produção mundial no ano 2022/2023, chegando a 168,2 milhões de sacos de café de 60kg, com o Brasil sendo o país com maior produção do grão, com 31,3% deste montante.<sup>7</sup>

A magnitude desse mercado se estende mundialmente com aproximadamente mais de 70 países no mundo cultivando o grão em suas terras, gerando aproximadamente 125 milhões de empregos formais e informais.<sup>2</sup> A Figura 1 mostra a média de produção de grãos de café, em 2018, ao redor do globo.

Figura 1 - Produção mundial de café 2022

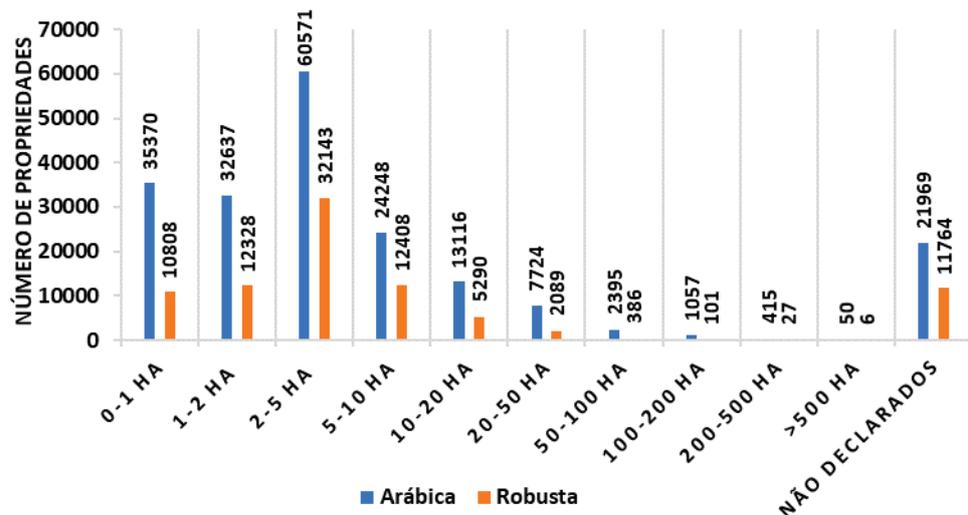


FONTE: Adaptado de <sup>8</sup>.

O café é o principal exportação de diversos países, como é o caso do Burundi, com cerca de 60% do seu mercado exportador focado no grão de café, Honduras 25% e Nicarágua 20%.<sup>2</sup> No Brasil, o tamanho do mercado cafeeiro reflete em aproximadamente 280 mil propriedades espalhadas em mais de 10 estados brasileiros gerando cerca de 8 milhões de empregos diretos e indiretos. Além disso, em 2012, cerca de 43% da produção anual foi destinada para o mercado interno, e em 2013 o consumo per capita chegou a 6,4 kg/habitante.

Um aspecto social importante na produção do café é que cerca de 85% das 280 mil propriedades responsáveis pelo plantio do café são pertencentes a pequenos e médios produtores, além disso, 76,86% das propriedades produtoras são de até 10 hectares.<sup>9</sup> A Figura 2 mostra a distribuição do número de propriedades de cultivo de café, separadas pelo tipo de café, por hectares, no Brasil.

Figura 2 –Propriedades produtoras de café do tipo Arábica e Robusta no Brasil.



FONTE: Adaptado de <sup>9</sup>, a partir de IBGE 2006.

## 2.2 CONTROLE DE QUALIDADE DO GRÃO DE CAFÉ

A grande demanda mundialmente de café faz com que a qualidade do produto seja um fator crítico para o seu sucesso. A associação Brasileira da Indústria de Café (ABIC), em seu regulamento técnico lista alguns parâmetros físico-químicos e seus limites máximos para caracterizar o produto<sup>10</sup>, mas tais valores e análises variam de acordo com a legislação do destino final do café.

Em um nível inicial, na fazenda, a qualidade está associada a produtividade, preço e facilidade do cultivo. Já para o consumidor, a qualidade está mais vinculada

a características organolépticas e o preço do produto.<sup>11</sup> Uma etapa que tem bastante destaque no processo produtivo é a torragem do grão de café, proporcionando diferentes características organolépticas e colorimétricas para o produto final.<sup>12</sup> Nesta etapa, o controle de qualidade é feito por mestre de torragem que determinam o grau de torra por avaliações organolépticas (sabor geral, aroma, acidez e corpo) e parâmetros físicos do processo como temperatura da torragem e tempo dentro do forno.<sup>13</sup>

Buscando garantir um produto com qualidade assegurada, a portaria N° 570, de 9 de maio de 2022 da secretária de defesa agropecuária (SDA), do ministério da agropecuária, pecuária e abastecimento (MAPA) que entrou em vigor no dia 1 de janeiro de 2023, tem como uma das suas disposições gerais estabelecer um padrão oficial de classificação do café torrado no território brasileiro para garantir segurança alimentar ao consumidor. O documento cria uma base legal para o controle de qualidade para cada tipo de grão do gênero *Coffea* e suas variações de processo produtivo.

Estão entre os parâmetros de qualidade aspectos sensoriais como fragrância do pó, aroma da bebida, acidez, amargor, sabor, sabor residual, corpo, percepção dos defeitos na bebida e adstringência a serem classificados para melhor direcionamento ao consumidor. Outro parâmetro para caracterizar o produto é o ponto da torrefação do café (clara, média e escura) que faz uso de uma classificação em função da cor do grão. Esse, apesar de não ter caráter probatório, caso haja divergência da informação da torra informada na embalagem com a torra real, é suspensa a distribuição do produto até a reembalagem com a informação correta.

E por último, com caráter desclassificatório, a portaria trás em anexo a Tabela 1, na qual quantifica os limites permitidos de matéria estranha (grãos ou sementes de outra espécie, corantes, açúcares, caramelo e borra de café solúvel ou de infusão) e/ou impurezas (cascas, paus e outros detritos proveniente do próprio cafeeiro) no café torrado.<sup>3</sup>

Tabela 1. PARÂMETROS DE QUALIDADE DO CAFÉ TORRADO

<b>Parâmetro</b>	<b>Tipo Único</b>
Somatório de matérias estranhas e impurezas	Máximo 1,0% (m/m) *
Elementos estranhos	Ausência**
Teor de cafeína no café descafeinado	Máximo 0,1%

\* Matéria estranha e impureza acima de 1,0% (um por cento) é considerado desclassificado

\*\* Café com presença de elementos estranhos é considerado desclassificado

FONTE: Adaptado de <sup>3</sup>.

Pela definição da norma, materiais estranhos são objetos encontrados no produto com origem desconhecida, podendo ser grãos de outras espécies, pedras ou terra. Impurezas são objetos que podem ter origem da lavoura do café, como casca de café, paus e grão de café.

A grande diferença entre matérias estranhas e impurezas para elementos estranhos é quando existe a constatação de uma intenção na adulteração, caracterizando o ato ilícito e desqualificando o produto.

Apesar da grande rigorosidade nos limites permitidos, a norma não apresenta nenhum método ou metodologia oficial de quantificação dos elementos estranhos na amostra de café torrado moído, apenas fazendo menção para métodos macro e microscópios, sem se aprofundar no tema.

Pesquisadores da Fundação Ezequiel Dias elaboraram o “Atlas de microscopia – Café torrado e moído”, no qual é proposto uma metodologia de identificação visual dos materiais estranhos presente no café. Em resumo, a amostra comercial de café torrado moído é misturada a uma solvente orgânico, como clorofórmio e levado a secar. Em seguida, o resíduo sólido é levado a um estereomicroscópio onde um especialista irá diferenciar, grão a grão, os grãos de café torrado moído de materiais estranhos e impurezas.<sup>4</sup>

Desse modo, a análise proposta é extremamente lenta e subjetiva, visto que é preciso um profissional extremamente capacitado para determinar a diferença entre grãos tão similares.

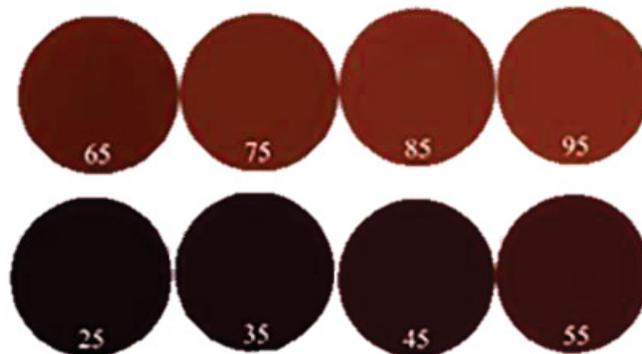
Com isso, outras metodologias mais robustas, sem subjetividade, veem sendo explorados afim de dar mais confiança a agilidade para a análise, assegurar a qualidade do produto final. Os próximos tópicos trazem alguma das ferramentas e como podem ser uma opção para garantir a qualidade do café de acordo com a legislação.

### 2.2.1 Espectroscopia ultravioleta-visível (Uv-VIS)

No controle de qualidade do café, a espectroscopia na região do UV-Vis pode ser usada na verificação do grau de torra do café. Esse parâmetro é extremamente importante visto que cada grau é responsável por uma experiência sensorial totalmente diferente para o consumidor, logo, controlar esse parâmetro garante o melhor destino para o café torrado moído.

Um dos métodos disponíveis para verificar o grau da torra sem a subjetividade que um mestre de torragem traz é utilizando o espectrofotômetro na região do visível e discos Agtron<sup>14</sup> da Associação Americana de Cafés Especiais (*Specialty Coffee Association of America* - SCAA) que padronizou a cor da torra. A escala vai de 0 a 100, onde o valor de 35 representa uma torragem mais escura, 55 torra média e 75 torra clara, como apresentado da Figura 3. Essa comparação pode ser feita visualmente com auxílio de discos comercializados pela SCAA ou com espectrofotômetros comercializados com tal função usando uma calibração na região do visível.<sup>15</sup>

Figura 3 –Discos de cor e escala numérica referente ao grau de torra definido pela SCA.



FONTE: Adaptado de<sup>16</sup>.

Os valores de referência Agtron são comumente usados em trabalhos envolvendo que buscam desenvolver alternativas mais baratas para classificação da torra. Lema D. S. usa a escala Agtron para obter valores de referência e correlacionar com valores dos canais de cores de uma imagem digital do grão, obtendo um valor de  $R^2$  de 0,99<sup>16</sup>, dando perspectiva de uma classificação confiável da torra com uma simples foto.

Lembrando que mais da metade dos produtos de café são de pequenos e médios produtos, o desenvolvimento de ferramentas simples e acessíveis possibilita um controle de qualidade rápido e confiável.

Outra possível aplicação da espectroscopia UV-Vis foi explorada por Souto U. T. C. P., que investigou por espectros UV-Vis (239 – 405nm) os efeitos em extratos do café quando adulterados com cascas e paus e desenvolveu um modelo classificatório usando algoritmo de projeção sucessiva associado com análise de discriminação linear (sigla em inglês, SPA-LDA) com valores de acurácia, sensibilidade e especificidade de 100%<sup>17</sup>, identificado a presença de adulterantes em amostras aquosas de café.

### 2.2.2 Espectroscopia infravermelho próximo (NIR)

A espectroscopia na região do infravermelho próximo (em inglês, near infrared, sigla, NIR) trabalha na faixa espectral de 780 a 2500 nm, e é uma potente ferramenta no controle de qualidade do café devido apresentar bandas vibracionais das componentes químicas que caracterizam este produto.<sup>13</sup> Com isso, diferentes variedades de café podem ser discriminadas em função das suas composições distintas. Baqueta M. R. explora em seu trabalho as diferenças entre as variedades de *Coffea canephora* torrado e moído e seus distintos produtores ao redor do Brasil com um infravermelho próximo (1500-2500 nm) e confirmou a potencialidade do método como possível ferramenta para rastrear e identificar a variedade e origem do produto, assegurando o registro de indicação geográfica que garante a procedência de um café de determinada região, agregando valor e segurança ao café.<sup>18</sup>

A avaliação dos parâmetros de qualidade pode ser também explorada em grãos de café verdes. Compostos como proteínas e lipídeos são comumente associados a qualidade final do produto. Todavia, sua determinação por métodos de referência, como o método Kjeldahl na determinação de proteínas, se mostra ambientalmente agressivo e demorado, principalmente em uma escala industrial.

Com isso, metodologias mais rápidas e não destrutivas podem garantir a segurança do produto de maneira mais eficiente. Zhu M em seu trabalho analisou amostras de café de diferentes origens por espectroscopia de infravermelho próximo na região de 830 a 2500 nm com forte absorção de bandas de combinação e sobretons das ligações químicas CH, NH, OH e SH. Com auxílio de ferramentas quimiométricas desenvolveu um modelo de calibração multivariada com coeficiente de determinação de 0,982 e 0,991 e limites de quantificação de 1,03g/100g e 0,46g/100g para lipídeos e proteínas, respectivamente.<sup>19</sup>

O grau de torra do grão de café também pode ser controlado nessa região espectral. A mudança visual que ocorre durante o processo de torrefação do café gera mudanças estruturais na composição do grão proporcionando diferenças espectrais.<sup>13</sup> Essa relação já é explorada por empresas que comercializam equipamentos que trabalham na região do infravermelho próximo e dão respostas na escala Agtron para determinação da torra.<sup>20</sup> Pires F. de C. também explorou essa correlação criando um modelo de calibração multivariada para avaliar a influência de diferentes graus de torrefação nos espectros do café torrado em grão e moído obtendo um coeficiente de correlação no modelo para as amostras de validação de 0,95 e 0,98, respectivamente.<sup>13</sup>

### 2.2.3 Imagens hiperespectral

A técnica de imageamento hiperespectral (em inglês, hyperspectral imaging, sigla, HSI) pode ser separado em duas partes: a primeira responsável pela aquisição da informação química e a segunda em cargo de vincular essa informação a um ponto do espaço, informação espacial.

Para adquirir a informação química geralmente se utilizam espectrofotômetros na região da radiação ultravioleta, visível, infravermelho próximo e médio, e Raman de modo a obter um espectro contínuo da amostra, com mais de 10 canais espectrais.<sup>21</sup> Para a parte espacial, utilizam-se mesas posicionadoras automatizadas responsáveis por assegurar que todos os pontos da amostra sejam analisados pelo espectrofotômetro.

Essa configuração proporciona um mapeamento das amostras sólidas em função das informações químicas que forem ativas na região espectral trabalhada, além de entender sua distribuição ao longo da área de analisada<sup>22</sup> gerando ao final do experimento um cubo de dados, de dimensões  $x$  e  $y$  que correspondem as coordenadas espaciais e o eixo  $z$  correspondente aos espectro. Esse cubo também é chamado comumente de “hipercubo”.<sup>23</sup> O tamanho desse cubo de dados depende de parâmetros experimentais como tamanho da área a ser analisada e a quantidade de pixels na imagem final (resolução espacial), e número de variáveis espectrais por pixel (resolução espectral).

As aplicações de técnicas de imageamento hiperespectral são inúmeras. A região espectral entre o final do ultravioleta e início do infravermelho próximo

(350nm-1050nm) já vem sendo explorada em diversas áreas como a biomédica, na identificação de células de câncer de mama com informação química entre 373-809nm<sup>24</sup> e na caracterização de pinturas e heranças culturais entre 375-1047nm.<sup>25</sup> Outra área que se beneficia fortemente dessa técnica é a indústria alimentícia buscando maior eficiência e praticidade. Chao K. fez o uso de HSI em tempo real em uma linha comercial de controle de qualidade de frangos de corte com velocidade análise de 140 aves por minuto para encontrar as variáveis de maior impacto na qualidade final do produto. O autor concluiu que o sistema tem grande potencial em verificação de aves com doenças sistêmicas.<sup>26</sup> Já Gowen A. desenvolveu uma metodologia promissora no controle de qualidade em função dos danos físicos encontrados em cogumelos brancos, aplicando PCA juntamente a imagens hiperspectrais entre 400 a 1000nm.<sup>27</sup>

A região do infravermelho próximo após 1000 nm também é bastante usado e aplicado de acordo com os componentes químicos de interesse. Tantantrakun A. buscou entender a maturação de abacaxi intactos com NIR-HSI na região de 935-1720nm. As frutas foram analisadas sem nenhum preparo, com casca, e obteve com um modelo de calibração com coeficiente de validação cruzada de 72% de sensibilidade.<sup>28</sup>

Buscando aplicações em amostra de café, Forchetti D. A. P. trouxe uma alternativa na detecção de adulterantes em café torrado e moído que pode ser uma alternativa para se adequar a portaria da SDA Nº 570 no controle de matérias estranhas, impurezas e elementos estranhos presentes no café torrado (Tabela 1) com imageamento hiperespectral na região do infravermelho próximo de 1282 nm a 2500 nm. Foram preparadas amostras de café com diferentes concentrações de diferentes adulterantes. Com auxílio de ferramentas quimiométricas foram analisados e identificados quimicamente e espacialmente a distribuição de adulterantes como casca, milho moído e torrado, paus e solo. Além disso os resultados foram promissores quanto a quantificação de adulterantes com erros menores que 4% (m/m).<sup>29</sup>

### 2.3 QUÍMIOMETRIA

Com o avanço na área instrumental na química e o desenvolvimento de equipamentos mais sofisticados, a geração de um grande volume de dados começou a fazer parte da rotina de trabalho dos pesquisadores. Um exemplo de

técnica que resulta em uma grande quantidade de informação em apenas alguns segundos são os equipamentos espectroscópicos, que registram centenas de valores de absorbância (ou intensidade) por comprimento de onda.<sup>30</sup> Tais experimentos podem gerar dados químicos, físicos e estatísticos, sendo o papel desse último atribuir confiabilidade ao experimento e extrair correlações entre os resultados, que de início, não são óbvias.

Entretanto, nem sempre a extração máxima de informação de um conjunto de dados é algo simples, além das informações experimentais importantes, como bandas de um espectro, há também fontes de variações aleatórias vinculadas ao equipamento, amostras comprometidas, variações de analista, etc. Tudo isso junto as informações relevantes do experimento, dificultam uma interpretação rápida e visual de comparações de dados manualmente. Como consequência, começou-se a desenvolver a área da quimiometria, um ramo interdisciplinar que envolve estatística, modelagem matemática e ciência da computação para aplicações químicas buscando uma forma de tratamento de dados sem subjetividade.<sup>31</sup>

Uma das principais áreas de atuação de métodos quimiométricos é em desenvolver modelos de calibração, validação e testes de significância.<sup>31</sup> E um dos ramos que se beneficia destes métodos é a indústria alimentícia, com métodos sendo gerados para controle de qualidade<sup>32,33</sup>, pesquisa e desenvolvimento de novos produtos, análises organolépticas vinculadas a dados experimentais<sup>34,35</sup>, e na área de processos de manufatura<sup>36</sup>.

### 2.3.1 Análise por componentes principais (PCA)

Em experimentos com um grande número de dados, convém se estudar a variação dos resultados de maneira eficiente, isso é, de modo a entender a máxima informação contida nos experimentos, ou seja, as informações mais relevantes. Tal abordagem, possibilita agrupar amostras em função da variância dos dados, uma vez que analitos parecidos tem respostas parecidas.

Um exemplo seria realizar uma análise exploratória em um conjunto de amostras de café arábica e robusta sem identificações por infravermelho próximo. Essa análise iria gerar um espectro varrendo todos os comprimentos de onda possibilitados pelo equipamento. Entretanto, tentar identificar diferenças espectrais visuais não é a maneira mais confiável para discernir entre ambas as variedades de

café. Logo, reduzir a dimensão dos dados sem perder a variação experimental facilita chegar em respostas de maneira mais simples e com resultados tão significativa quanto usando todas as variáveis, possibilitando o agrupamento das variedades em função das variações espectrais.

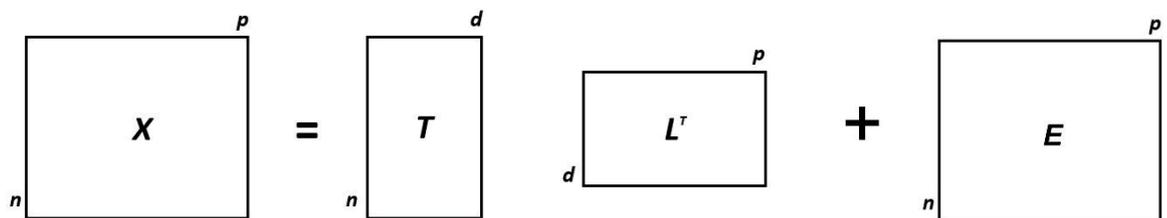
Assim, um método estatístico bastante usado na área de quimiometria é a análise por componentes principais (do inglês, *Principal Components Analysis* – PCA) que busca transformar as variáveis originais em um outro conjunto de variáveis ortogonais, chamadas de componentes principais, acumulando maior variância dos dados em um número reduzido de variáveis.<sup>31</sup>

Os dados de um experimento podem ser visualizados como uma matriz **X** contendo  $n$  linhas ( $n=1, \dots, N$ ; associado ao número de amostras) por  $p$  colunas ( $p=1, \dots, P$ ; variáveis) resultando em uma matriz de dimensões  $n \times p$ . Tendo em vista o foco da PCA em agrupar as variáveis ortogonais em componentes principais, este novo conjunto explica o conjunto de dados original com um número menor de variáveis, ocorre uma redução das dimensões iniciais do conjunto de dados, a matriz **X** pode ser reescrita como um produto de duas matrizes menores, mais o seu respectivo resíduo, como apresentado nas Equações 1 e 2:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (1)$$

onde **T** é a matriz de escores de  $n$  linhas e  $d$  colunas (número de componentes principais), **P<sup>T</sup>** é a matriz de pesos transposta com  $d$  linhas e  $p$  colunas, e **E** são os resíduos, com mesma dimensão da matriz **X**, que carrega a informação das demais componentes principais não utilizadas. A Figura 4, mostra a representação matricial dessa decomposição.

Figura 4 - Redução das dimensões dos dados iniciais.



FONTE: Adaptado de <sup>37</sup>.

A contribuição de cada termo pode ser calculada pela equação abaixo

$$\text{Porcentagem de Variância Explicada}(\%) = 100 \frac{\sum_{n=1}^N \sum_{p=1}^P (t_{np})^2}{\sum_{n=1}^N \sum_{p=1}^P x_{np}^2} \quad (2)$$

O número de componentes principais pode ser igual ao número de variáveis, todavia, a redução dimensional ocorre quando algumas componentes já representam uma porcentagem de variância dos dados acima do seu ruído instrumental, conseqüentemente, as demais componentes podem ser ignoradas, ocorrendo assim uma diminuição na dimensão dos dados. Ghosh D. faz o uso dessa ferramenta a fim de identificar parâmetros sensoriais críticos para a aprovação de produtos alimentícios fermentados e concluindo a eficiência do método em direcionar para as variáveis que mais explicavam as variâncias sensoriais.<sup>38</sup>

### 2.3.2 Mínimos quadrados parciais (PLS)

A técnica de mínimos quadrados parciais (do inglês, *Partial least squares* - PLS) foi inicialmente desenvolvido por H. Wold no final da década de 60 para aplicação no campo da econometria. A popularização do método na química foi feita pelos grupos liderados por S. Wold e H. Martens ao fim da década de 70<sup>31</sup>. O método consiste em buscar uma correlação entre as variáveis **Y** e as variáveis exploratórias **X**, gerando novas variáveis, denominadas variáveis latentes. Conseqüentemente, as novas variáveis buscam concentrar o máximo de informação proveniente dos dados contidos em **X** e **Y**, que foram usados para sua obtenção.<sup>39</sup> A principal diferença para o método de PCA é a utilização das informações das variáveis independentes da matriz **Y** em sua decomposição matricial.

A equação 3 apresenta como essa ferramenta associa um método de referência, variáveis **Y**, a um método rápido, variáveis **X**. Considere **Y** sendo uma matriz com resultados químicos de referência e **X** a mesma amostra analisada em **Y**, so que em um equipamento rápido e de baixo custo

$$Y = X b \quad (3)$$

A ferramenta PLS irá solucionar essa equação para o vetor de regressão, **b**, possibilitando realizar previsões de novos valores de **Y**, com base apenas no experimento de **X**.

O trabalho de Zhu M, apresentado em tópicos anteriores, criou uma metodologia de previsão de proteínas e lipídeos em amostras de café com um equipamento de infravermelho próximo usando como referência o método Kjeldahl. Nesse exemplo, a matriz X é o resultado espectroscópico e a matriz Y o método de referência. Com isso, aplicando PLS, obteve resultados com erro menores de 0,102 mg/g.<sup>19</sup>

Para avaliar o modelo de calibração, os parâmetros: raiz quadrada do erro médio quadrático de validação cruzada (RMSECV, do inglês *root mean square error of cross-validated*), raiz quadrada do erro médio quadrático de previsão (RMSEP, do inglês *root mean square error of prediction*) e raiz quadrada do erro médio quadrático de calibração (RMSEC, do inglês *root mean square error of calibration*) indicam se o modelo de calibração é eficiente. Outros parâmetros que também podem ser úteis é o coeficiente de determinação de validação cruzada ( $R^2_{cv}$ ) e de previsão ( $R^2_{pred}$ ). Segue as respectivas equações dos parâmetros:

$$RMSEP = \sqrt{\frac{1}{m} \sum_{j=1}^m (\hat{y}_{Pj} - y_{Pj})^2} \quad 5)$$

$$RMSEC = \sqrt{\frac{1}{n} \sum_{i=1}^m (\hat{y}_{Ci} - y_i)^2} \quad 6)$$

$$R^2_{CV} = 1 - \frac{\sum_{i=1}^n (\hat{y}_{CVi} - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad 7)$$

$$R^2_{Pred} = 1 - \frac{\sum_{j=1}^m (\hat{y}_{Pj} - y_j)^2}{\sum_{j=1}^m (y_j - \bar{y})^2} \quad 8)$$

Onde  $y_i$  é o valor de resposta da amostra  $i$ ,  $\hat{y}_i$  é o valor de resposta predito pelo modelo da amostra  $i$  (considerando o subíndice  $C$  para o modelo durante a etapa de calibração,  $CV$  para o modelo durante a etapa de validação cruzada e  $P$  para o modelo durante a etapa de previsão) e  $\bar{y}$  é o valor médio das respostas.

### 2.3.2.1 Mínimos quadrados parciais com análise discriminantes (PLS-DA)

A análise de mínimos quadrados parciais com análise discriminante (do inglês, *Partial least squares Discriminant Analysis* – PLS-DA) é uma variação da PLS que tem como objetivo separar as amostras em grupos de categorias/classes.

A diferença entre esse método para o PLS tradicional é a presença de variáveis independentes, na matriz **Y**, do tipo categórica (*dummy variables*, em tradução livre variáveis burras), usando valores binários, 0 ou 1, que indicam se a amostra de interesse tem determinada característica ou está inserida em determinada classe.<sup>40</sup>

Para avaliação do modelo, são usadas as figuras de mérito (sigla, FoM) relacionada a quantidade de predições corretas e incorretas para cada grupo de interesse. A tabela de confusão ou tabela de contingência (Tabela 2) dá panorama da quantidade de atribuições corretas de cada grupo modelado.

A quantidade de amostras do grupo 1 preditas corretamente como pertencente ao grupo 1 é chamado de casos verdadeiro positivo (sigla, VP). Caso ocorra atribuição errônea de uma amostra do grupo 1 ao grupo 2, temos a situação de um Falso negativo (sigla, FN). Analisando o grupo 2, a predição correta dessa amostra é definida como casos Verdadeiros negativos, e para os casos errados, Falso positivo.

Tabela 2 – Tabela de confusão a partir de um modelo PLS-DA genérico.

Tabela de confusão		Grupos reais	
		GRUPO 1	GRUPO 2
Predito pelo MODELO	Predito como GRUPO 1	VP	FN
	Predito como GRUPO 2	FP	VN

FONTE: adaptado de <sup>41</sup>.

Com a Tabela 2 podemos calcular a taxa de acerto e taxa de erro para cada grupo modelado chamados de sensibilidade e especificidade, respectivamente, para cada grupo. A sensibilidade do grupo 1 é definida como a taxa de predições de amostras do grupo 1 ao grupo 1 e especificidade é taxa de atribuições do grupo 2 ao grupo 1. Segue abaixo as formulas para cada uma destas figuras de mérito

$$\text{Sensibilidade}(\text{grupo } x) = \frac{VP(\text{grupo } x)}{VP(\text{grupo } x) + FP(\text{grupo } x)} \quad 8)$$

$$\text{Especificidade}(\text{grupo } x) = \frac{VP(\text{grupo } y)}{VP(\text{grupo } y) + FP(\text{grupo } y)} \quad 9)$$

$$\text{Eficiência} = \frac{VP(\text{grupo } x) + VP(\text{grupo } y)}{VP(\text{grupo } x) + VP(\text{grupo } y) + FP(\text{grupo } x) + FP(\text{grupo } y)} \quad 10)$$

Suhandy D. faz o uso dessa ferramenta quimiométrica juntamente com espectros UV-Vis para o categorizar os grãos de café entre duas espécies de café robusta alcançando uma classificação de 100% dos grãos corretamente <sup>42</sup>. Uma aplicação similar feita pelos mesmos autores foi no monitoramento de adulteração no Kopi Luwak (Café Civeta) com auxílio de espectros UV-Vis, alcançando uma classificação de 100% das amostras.<sup>43</sup> Outra possibilidade dessa ferramenta é identificar padrões regionais nas amostras de interesse, Bassbasi M. identificou diferenças entre manteigas de diferentes partes de Marrocos por infravermelho por transformada de Fourier (FTIR) combinados com PLS-DA, obtendo uma precisão de 100%.<sup>44</sup>

### 2.3.3 Data driven-soft independent modeling of class analogy (DD-SIMCA)

Em problemas reais de autenticidade, é desejável constatar se uma amostra está de acordo com as características físicas e/ou químicas de um grupo referência/alvo, não importando com o que foi adulterado. Para tanto é necessário apenas ter conhecimento das informações físicas e/ou químicas do grupo que deseja controlar. Métodos de classificação de uma classe (em inglês, *one-class classification*, sigla, OCC) buscam criar modelos capazes de predizer se uma amostra desconhecida está dentro do grupo alvo ou não, não importando qual seja as características que excluem a amostra do grupo de interesse. Desta forma, só é necessário ter conhecimento apenas do grupo de referência, extinguindo assim a necessidade de banco de dados de adulterantes existentes e os novos que continuamente são criados.<sup>45</sup> Entre os métodos de OCC, a ferramenta *Data-driven*

*Soft Independent Modeling of Class Analogy*, sigla DD-SIMCA, possibilita construir modelos com base nas informações do grupo alvo e estipular limites para predição da amostra desconhecidas, de forma a considerar se a nova amostra analisada faz parte do grupo alvo ou se é algo desconhecido.

Inicialmente esta ferramenta faz o uso da decomposição da matriz  $\mathbf{X}(i \times j)$  por PCA, equação 1. A partir disso, dois parâmetros podem ser extraídos: a distância quadrada de Mahalanobis ou distância dos escores (score distance, sigla, SD),  $h_i$ , que é referente a posição de uma amostra  $\mathbf{x}$  ( $1 \times j$ ) no subespaço dos escores definidos pelas  $d$  componentes principais; e a distância ortogonal (Orthogonal distance, sigla, OD),  $v_i$ , relacionada a soma quadrática dos resíduos extraída da matriz  $\mathbf{E}$ , associada a distância da amostra  $\mathbf{x}$  em relação ao subespaço das  $d$  PCs<sup>46,47</sup>. Ambos os parâmetros normalizados por sua respectiva média podem ser aproximados por uma distribuição chi-quadrado escalada:

$$N_h \frac{h_i}{h_0} \sim \chi^2(N_h) \quad (12)$$

$$N_v \frac{v_i}{v_0} \sim \chi^2(N_v) \quad (13)$$

sendo  $N$  o grau de liberdade de cada um dos parâmetros.

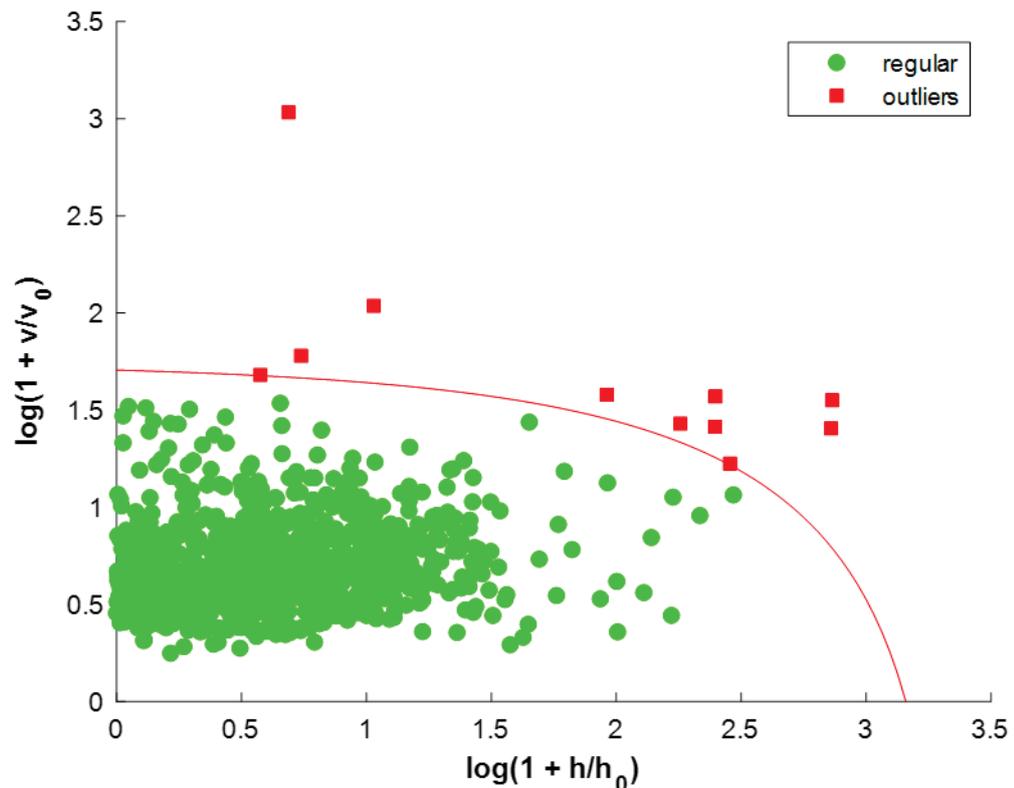
Posto isso, é possível construir uma área de aceitação em função de ambas as distribuições que por sua vez estão ligadas as informações que caracterizam o grupo alvo ao estipular um valor  $\alpha$ , definido como erro  $\alpha$  (Tipo I) para criar uma área de aceitação descrita por:

$$c_{crit} = \chi^{-2}(1 - \alpha, N_h + N_v) \quad (14)$$

$$c \leq c_{crit} \quad (15)$$

Com isso, se uma amostra pontuar maior do que  $c_{crit}$ , é considerada como fora do grupo alvo. A Figura 5 mostra a área de aceitação que é construída a partir da distribuição chi-quadrado dos parâmetros  $h_i$  e  $v_i$ .  $h_0$  e  $v_0$  são as respectivas médias de cada termo.

Figura 5 - Gráfico de aceitação DD-SIMCA. Círculos verdes são amostras dentro do grupo alvo e quadrados vermelhos são amostras de um grupo alternativo.



FONTE: próprio AUTOR (2023).

Por fim, o modelo pode ser avaliado a partir da sensibilidade dada pela equação 8, especificidade só pode ser calculada quando é conhecida a classe alternativa.<sup>47,48</sup>

As aplicações de autenticação têm grande espaço na área alimentícia. Souza R. R. de explorou aplicação do DD-SIMCA em 73 amostras de grupo alvo, mel puro, e 162 amostras do alvo adulteradas com xarope de milho, xarope de agave e/ou melaço de cana-de-açúcar em soluções aquosas no espectrômetro UV-Vis entre 200-800nm e obteve sensibilidade 96% com simples pré-tratamento <sup>49</sup>. Neves M. de G. mostrou a possibilidade do controle da autenticidade de óleo de coco virgem a partir de um equipamento de ATR/FTIR obtendo 100% de sensibilidade, além disso, explorou amostras adulteradas com óleo de milho, girassol, canola e soja entre 10 a 40% afim de indicar a composição da adulteração e obteve valores de sensibilidade entre 88% a 100% para os respectivos modelos dos adulterantes. Estudos em amostras sólidas de gorduras de carne de porco por Totaro M. P. investigou sistemas de criação distintos em função da alimentação das

suiniculturas com espectroscopia NIR e obteve uma sensibilidade de 100% solucionando o problema de autenticidade.<sup>50</sup> Por fim, estudos envolvendo a origem do café de Santos L. B. dos alcançou 100% de sensibilidade com extratos de grão verde de café do cerrado mineiro, qual tem registro de denominação de origem protegida com um equipamento de espectroscopia Uv-VIS com a janela de trabalho entre 230-450nm.<sup>51</sup>

#### 2.3.4 Resolução de curva multivariada por mínimo quadrados alternados (MCR-ALS)

Os resultados obtidos pelo imageamento espectral na região do infravermelho próximo no trabalho de Forchetti D. A. P possibilitou mapear uma área da amostra de café misturado com adulterantes e encontrar informações químicas distintas para o café e do adulterante e suas respectivas distribuições no espaço analisado.<sup>29</sup> Cada pixel tem uma posição no espaço e associado a isso um espectro resultado da mistura das componentes químicas que estão presentes nele. No caso de Forchetti D. A. P cada pixel tem diferentes proporções da mistura de café e um adulterante.<sup>29</sup> Portanto um espectro é a soma dos espectros das componentes puras, café e adulterante, distribuídas na amostra em função da sua respectiva concentração, respeitando a lei de Lambert-Beer que pode ser representada pelo modelo bilinear.<sup>21</sup>

Inicialmente uma das primeiras ferramentas matemáticas aplicadas na tratativa dos dados de HSI foi análise de componentes principais (PCA) devido sua formulação (Equação 1) ser um modelo bilinear. Todavia os resultados obtidos são restritos a informações não correlacionadas, isso é, ortogonais entre si, algo que dificilmente ocorre quando estamos tratando de componentes químicas que comumente tem informações repetidas.<sup>21</sup> Além disso, os resultados de *scores* e *loadings* não contribuem quimicamente para a discussão de maneira direta, um perfil de *loadings* não pode ser comparado a um perfil espectral. Portanto a PCA não é a ferramenta mais adequada para extração máxima dos resultados de HSI.

A resolução de curvas multivariada por mínimos quadrados alternados (em inglês, *Multivariate curve resolution-alternating least squares*, sigla, MCR-ALS) apresenta um modelo bilinear:

$$D = CS^T + E \quad (11)$$

onde descreve o conjunto de dados do imageamento **D** como o produto do perfil de concentração **C** multiplicado pelo perfil espectral puro das componentes **S** mais a variância não explicada pelo modelo **E**. Essa ferramenta substitui a necessidade de requisitos como ortogonalidade por restrições, que fazem mais sentido em dados de origem química. Algumas dessas restrições são: a não negatividade para o perfil de concentração e espectral, unimodalidade quando se assume um único máximo para perfil a qual foi aplicado, balanço de massa entre as espécies, correspondência entre as espécies do sistema, etc. Além disso, para dar início aos cálculos é necessário uma estimativa inicial para o perfil de concentração **C** ou de espectro **S**. Para tanto, existe a possibilidade de fornecer o perfil puro caso já seja conhecido ou obter o mesmo por análise multivariada por auto modelagem interativa (do inglês, Simple-to-use interactive self modeling multivariate analysis, sigla, SIMPLISMA) qual faz uso de ferramentas básicas de estatística para propor a separação inicial de sinal<sup>52</sup> ou a ferramenta de análise de fatores evolucionária (do inglês, evolving factor analysis, sigla EFA) segmentando a matriz e avaliando os possíveis postos (*ranks*) químicos.<sup>53</sup>

Uma vez estimado **C** e **S**, as restrições atribuídas são exploradas iterativamente com a resolução de quadrados mínimos alternados, Equações 12 e 13, buscando otimizar as os perfis recuperados

$$\mathbf{C} = \mathbf{D}\mathbf{S}(\mathbf{S}^T\mathbf{S})^{-1} \quad (12)$$

$$\mathbf{S}^T = (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{D} \quad (13)$$

Por fim, o modelo gerado faz uso de duas figuras de mérito para avaliar a decomposição espectral obtida. A falta de ajuste (em inglês, lack of fit, sigla, lof) entre a matriz decomposta e a matriz original, e a percentagem da variância explicada pelo modelo  $R^2$ , Equação 14 e 15 respectivamente.

$$\text{lof} = 100 \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^m e_{ij}^2}{\sum_{i=1}^n \sum_{j=1}^m x_{ij}^2}}, \quad e_{ij} = x_{ij} - \hat{x}_{ij} \quad (14)$$

$$R^2 = 100 \frac{\sum_{i=1}^n \sum_{j=1}^m x_{ij}^2 - \sum_{i=1}^n \sum_{j=1}^m e_{ij}^2}{\sum_{i=1}^n \sum_{j=1}^m x_{ij}^2}, \quad (15)$$

Com isso é possível separar e atribuir a distribuição espacial de diferentes componentes químicos em um imageamento hiperespectral como foi feito no trabalho de Forchetti D. A. P.<sup>29</sup> Outra aplicação dessa ferramenta presente na literatura é a adulteração de amendoim em chocolate em pó, Laborde A. realizou um mapeamento das amostras e identificou quimicamente, e espacialmente, o adulterante utilizando MCR-ALS<sup>54</sup>. Além da aplicação em imageamento de alimentos, a área forense, mais especificamente da documentoscopia, pode fazer uso de MCR-ALS. Borba F. de S. L. identificou rasuras em folha de papel e conseguiu recuperar a informação inicial, que foi censurada, com imageamento com Raman cofocal com MCR-ALS aplicado ao conjunto de dados.<sup>55</sup>

Sendo assim, esse trabalho busca preencher a lacuna presente na garantia da qualidade de café torrado moído de maneira rápida e sem subjetividade, identificado se existe a presença de impurezas e/ou elementos estranhos na amostra. Para isso, um equipamento de imageamento hiperespectral de baixo custo e diferentes ferramentas quimiométricas serão aplicados para avaliação de amostras de café torrado moído adulterados com casca de café, milho, cevada, caroço de açaí e trigoilho.

### **3 OBJETIVOS GERAIS E ESPECÍFICOS**

#### **3.1 OBJETIVO GERAL**

Desenvolver uma metodologia analítica de baixo custo baseada na técnica de imageamento hiperespectral capaz de identificar a presença do tipo de café e possíveis adulterantes em amostras de café torrado e moído.

#### **3.2 OBJETIVO ESPECÍFICO**

Classificar a espécie de café presente em amostras adulteradas usando ferramentas espectroscópicas

Detectar a presença de adulterantes em amostras de café usando espectroscopia na região do infravermelho próximo

Detectar a presença de adulterantes em amostras de café usando imageamento hiperespectral na região do visível

Com auxílio da ferramenta quimiométrica MCR-ALS extrair as componentes puras que constituem a amostra e sua distribuição espacial.

Identificar as componentes químicas presentes no analito e comparar com seus respectivos espectros puros com ferramentas estatísticas de validação.

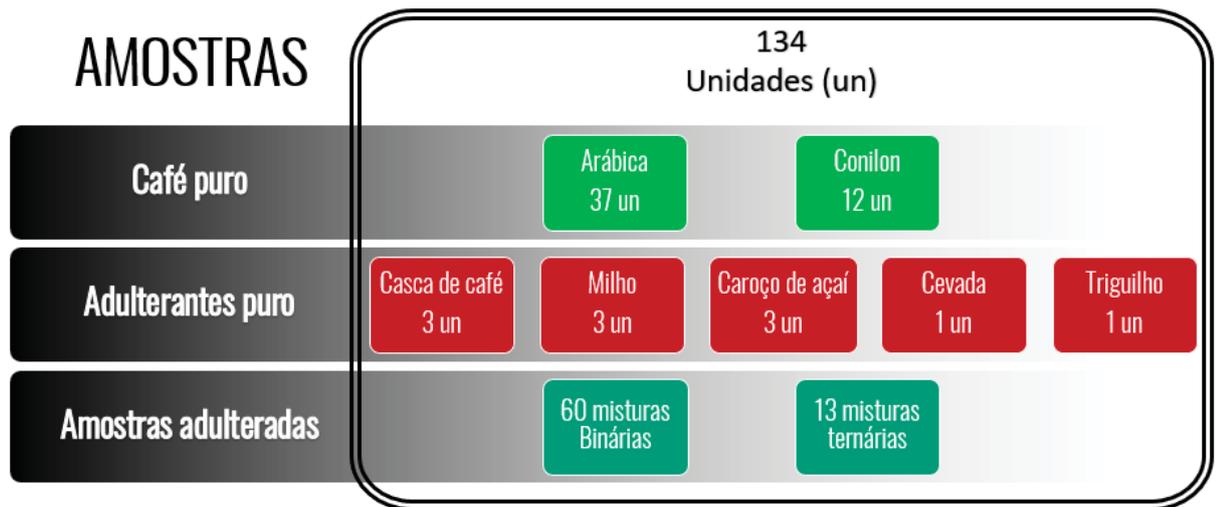
## 4 MATERIAIS E MÉTODOS

Neste capítulo serão listados os materiais necessários para realização do projeto e os passos metodológicos para preparo das amostras até aplicação dos métodos quimiométricos.

### 1. Obtenção das amostras

Foram enviadas pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA) 134 amostras torrada e moída, nas mesmas condições de torrefação e moagem, de café e adulterantes (Figura 6). É importante ressaltar que não houve nenhum tratamento dessas amostras. O objetivo é simular amostras reais, o mais similar possível a uma amostra de café torrado moído pronto para o consumo humano.

Figura 6 - Amostras enviadas pelo MAPA.



FONTE: O autor (2024).

O grupo de amostras consiste em: 37 e 12 unidades de café arábica e café conilon puros, respectivamente, 3 unidades de cada um dos adulterantes puros: casca de café, milho e caroço de açaí, 2 de triguilho e 1 de cevada, e 60 amostras de mistura binária de um tipo de café mais um adulterante, e 13 de misturas ternárias composta pelas duas espécies de café e um adulterante com concentração de adulteração entre 3 a 20%(m/m).

### 2. Espectros NIR

Todas as amostras foram analisadas utilizando um espectrofotômetro NIR da PerkinElmer, modelo Spectrum Two, com acessório de reflectância NIR, e resolução de 16 nm. As amostras foram homogeneizadas e acomodadas em placas de Petri para realização das leituras no equipamento. Os espectros foram adquiridos em triplicatas reais, isso é, um espectro para cada mistura de mesma composição, para cada amostra na faixa de 800 a 2750 nm. Estas análises foram realizadas pelo Laboratórios Federais de Defesa Agropecuária do MAPA e cedidos para o tratamento de dados desta dissertação de mestrado.

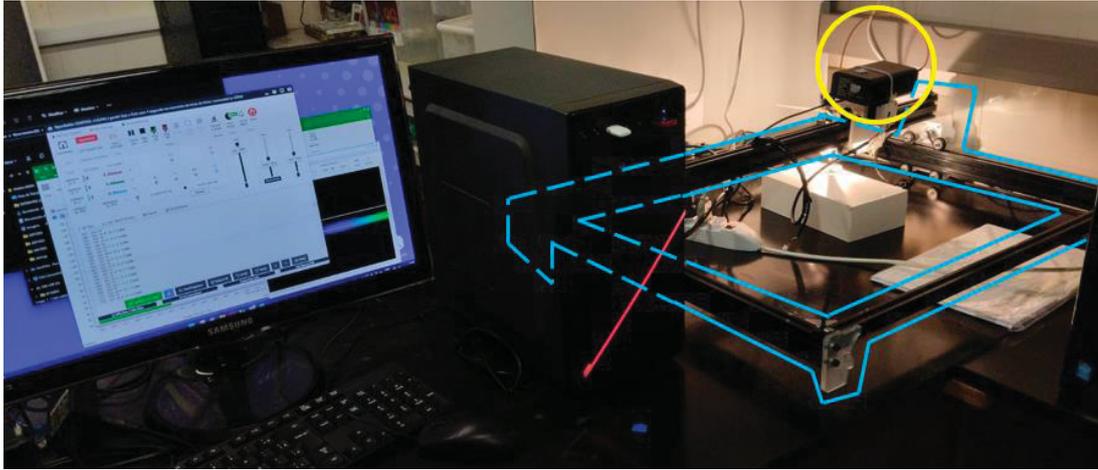
### 3. Espectros UV-vis-NIR

O espectrômetro utilizado neste trabalho foi o equipamento SMA-E, da ThuderOptics, com faixa de trabalho de 350 a 920nm e resolução de 0,29nm, contendo 1920 canais espectrais, acoplado a uma fibra optica da ThorLabs, modelo SMA905. Como fonte de radiação para aquisição do espectro na região do visível, utilizou-se uma lâmpada halógenas de alta intensidade luminosa (100W e OSRAM). Para controle do equipamento, foi utilizado o software, gratuito, SpectraGryph 1,2. As análises envolvendo esse equipamento foram feitas no Departamento de Química da Universidade Federal do Paraná, no laboratório DataLAB.

### 4. Imagens hiperespectral

Para aquisição da informação química, foi utilizado o espectrômetro detalhado na seção 3.3. A fibra óptica do espectrômetro foi posicionada em um ângulo normal a da superfície da amostra em uma distância de cerca de 0,2 mm. Para a informação espacial, foi utilizado uma mesa de controle numérico por computador (em inglês, computer numeric control, sigla, CNC), na qual o espectrômetro foi acoplado e fixado de maneira mecânica encima da parte móvel da mesa (Figura 7). A mesa foi controlada pelo software gratuito OpenBuilds v1.0.354, onde é programado o mapeamento das amostras.

Figura 7 - Espectrômetro (círculo amarelo) acoplado a mesa CNC (azul).



FONTE: próprio AUTOR.

O controle do espectrômetro e da mesa CNC foram feitas dentro de cada do seus respectivo ambiente. As componentes do equipamento de imageamento hiperespectral está descrito na Figura 8.

Figura 8 - Configuração do equipamento de imageamento hiperespectral de baixo custo desenvolvido no projeto de mestrado.

**CONFIGURAÇÃO - IMAGEAMENTO HIPERESPECTRAL**

**INFORMAÇÃO QUÍMICA**

**Espectros UV-Vis-NIR**

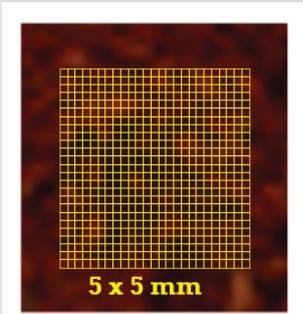
- ❖ Modo : Reflectância
- ❖ 1 espectro / 2 segundos

**INFORMAÇÃO ESPACIAL**

**Mesa CNC**

- ❖ Passo = 0,2 mm
- ❖ 1 passo / 2 segundos
- ❖ Intervalo entre passos = 2 segundos
- ❖ Resolução = 0,2 mm/pixel

**Espaço mapeado**



**5 x 5 mm**

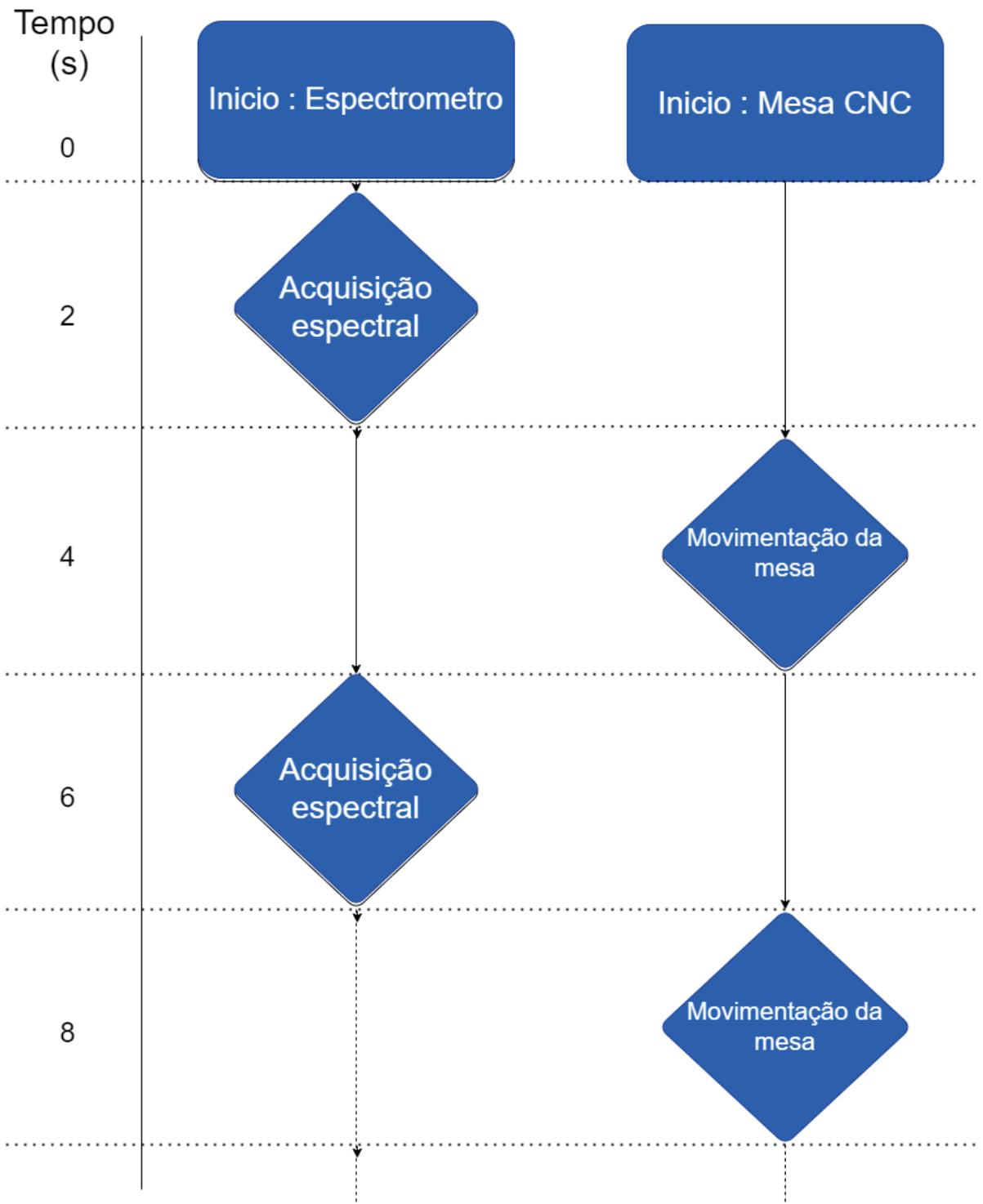
700 espectros / amostras  
Tempo de análise : 12 min

<R\$ 2500

FONTE: O autor (2024).

O espectrômetro foi configurado para realizar aquisições de espectros a cada 2 segundos no modo refletância. A mesa CNC também usou a mesma janela de tempo entre seu reposicionamento de 0,2mm. A Figura 9 mostra a ordem de execução de cada um dos softwares envolvidos no processo de imageamento.

Figura 9 - Ordem de execução do equipamento de imageamento hiperespectral.



Fonte: O autor (2024).

Com o custo final de menos de R\$ 2.500,00, considerando o espectrômetro UV-Vis e mesa CNC, é possível obter imagens com resolução de 0,2mm / pixel em uma análise de menos de 15 minutos.

### 5. Análises quimiométricas

Os dados foram analisados usando o software MATLAB versão R2019a com as ferramentas PLS\_Toolbox (para execução do PLS-DA e PCA), MCR-ALS GUI 2.0<sup>56</sup>, DD-SIMCA<sup>45</sup>, além de rotinas e funções desenvolvidas no próprio grupo de pesquisa.

Para todos os tratamentos quimiométricos com o espectro visível e espectros NIR cedidos pelo MAPA, foram utilizados apenas a região espectral de 367,30 a 749,05nm e 1126,8 a 2613,7nm respectivamente, devido grande ruído constatado no começo e fim de toda das regiões de trabalho dos equipamentos.

Para criação dos modelos dos tipos de café, visando uma performance rápida e confiável, dos 20300 espectros puros de café arábica (29 amostras) e 7700 de café conilon puros (11 amostras), foram selecionados de maneira aleatória 6720 espectros de cada uma das espécies. Para os espectros puros dos adulterantes, todos os espectros quando parecidos foram usados para calibração dos seus respectivos modelos e serão especificados quando contrário ao longo do texto.

Para classificação de cada espécie, os modelos PLS-DA foram criados usando o software PLS\_Toolbox. O método utilizado para validação cruzada foi venetian-blind (janela = 7) e a seleção do número de variáveis latentes foi considerando o menor valor atingido pelas equações 4 e 6 antes da estabilização em função do número de variáveis latentes. Os valores de sensibilidade e especificidade foram calculados pelas equações 8 e 9 respectivamente.

Para os testes de autenticidade, os modelos DD-SIMCA foram criados utilizando duas componentes principais, erro alfa  $\alpha$  (Tipo I) definido como 0,05 e área de aceitação estimada pela distribuição chi-quadrado. Para os conjuntos de dados usados na calibração dos modelos de adulterantes, foram utilizados todos os espectros puros. Os valores de sensibilidade de cada modelo foram calculados de acordo com a equação 8.

Para cálculo da estimativa inicial usado no MCR-ALS, foi usado o algoritmo SIMPLISMA. Entre as restrições a serem aplicados na otimização por mínimos quadrados alternados durante o MCR, foi usado não negatividade pelo algoritmo

fnnls (*fast non-negativity least squares*) para o cálculo do perfil recuperado de concentração para todos os conjuntos de dados, independente do pré-tratamento. Não-negatividade por fnnls foi usado para todos os pré-tratamento, exceto quando aplicado SNV. Os valores de falta de ajuste e variância explicada foram calculados pelas equações 14 e 15.

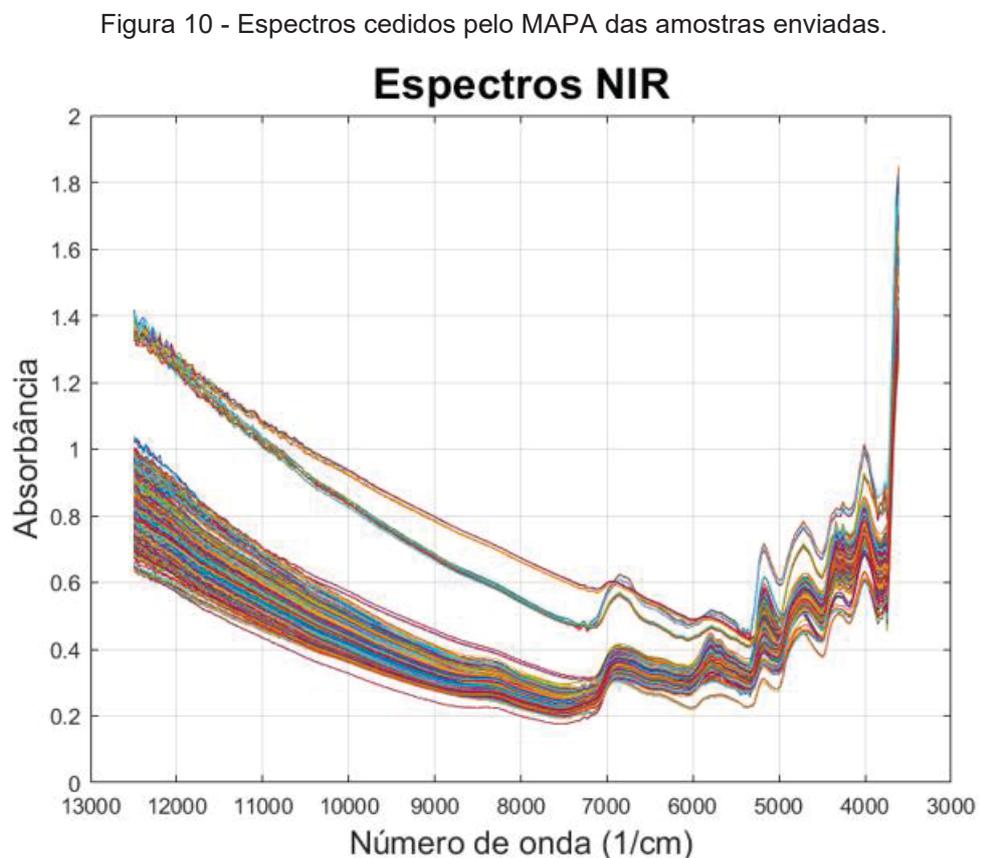
Para comparação entre os espectros puros e os recuperados pelo MCR-ALS, foi empregado coeficiente de correlação entre espectros  $r^2$ , juntamente de um filtro para evitar a dupla atribuição de uma componente recuperada por MCR-ALS ao mesmo tipo de espectro puro para as misturas binárias.

## 5 RESULTADOS E DISCUSSÃO

### 5.1 MODELO COM ESPECTROS NIR CEDIDOS PELO MAPA

Foram adquiridos um total de 490 espectros NIR das amostras, variando o número de replicadas de cada (Figura 10). Tendo a portaria Nº 570 como orientação para o desenvolvimento dos modelos, foi elaborado um modelo PLS-DA para a discriminação de dois grupos: café puro e café adulterado com níveis de 3,4,5,10 e 20% dos adulterantes citados na metodologia.

O pré-tratamento de escolha foi normalização por MSC com espectros mediano, derivada de 1ª ordem e centragem na média. Para o PLS-DA foram utilizadas 5 variáveis latentes, escolhidos a partir do menor erro de classificação obtido durante a etapa de validação cruzada.



FONTE: próprio AUTOR.

A tabela de confusão (Tabela 3) resume o desempenho do modelo para calibração e validação cruzada por venetian-blind (janela = 7). Foi obtida uma sensibilidade para café e adulterantes <95 %. Para validação externa do modelo, dois conjuntos de amostras foram usados: o primeiro conjunto das amostras foi selecionado realizando um kernnard-stone no conjunto das amostras previamente separadas para a calibração do modelo (metodologia I), o segundo conjunto continha amostras selecionadas pelo MAPA para validação do modelo. O segundo conjunto apresentava amostras de café diferentes daquelas usadas na construção das 134 amostras iniciais, com cafés de variedades e torrefações diferentes (metodologia II).

Tal abordagem foi realizada para utilizar o primeiro conjunto como forma de validação dos parâmetros de qualidade do modelo e possíveis sobreajustes, enquanto o segundo conjunto foi selecionado para testar a robustez do modelo frente a variedades e modos de preparo para os cafés diferentes. Com isso, a Tabela 3 apresenta os valores de sensibilidade e especificidade para validação externa insuficientes e não compatíveis com os resultados da validação cruzada.

Tabela 3 - Tabela de confusão do modelo PLS-DA para verificar a adulteração de café na região do NIR.

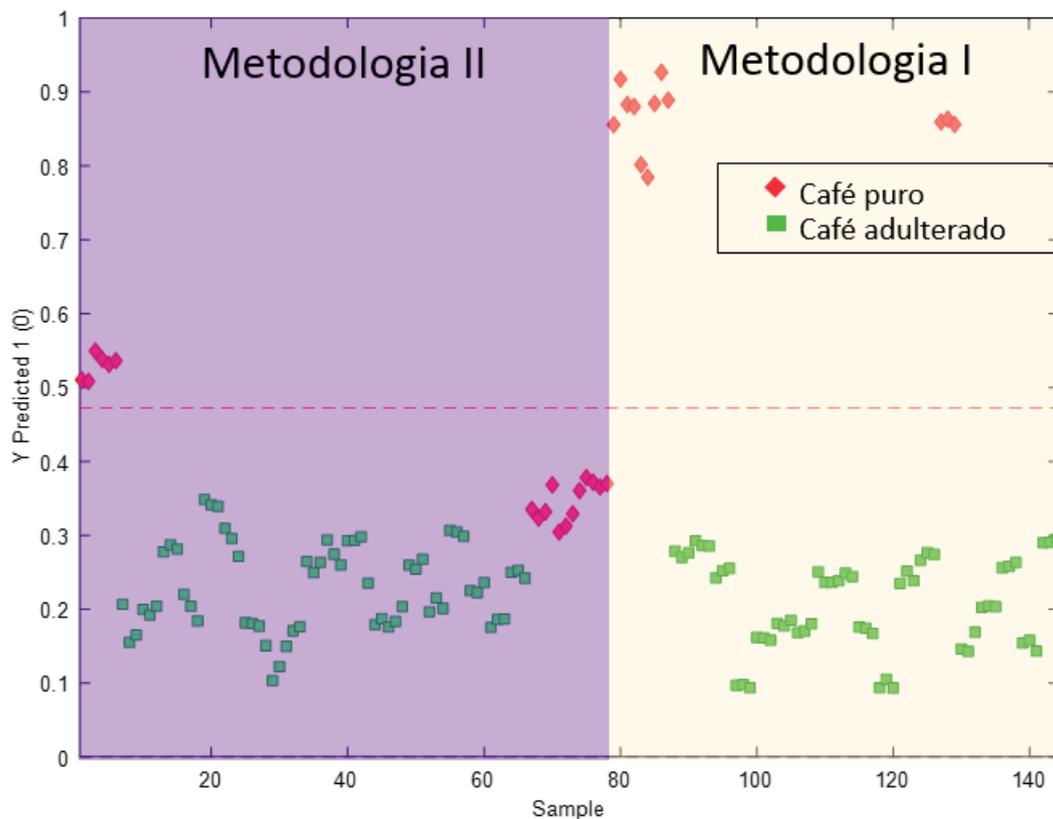
Tabela de confusão ( <b>Calibração e Validação cruzada</b> )				
Classes reais				
	Café puro	Café Adulterado	Sensibilidade	Especificidade
Predito como café puro	146	3	0,9669	0,9846
Predito como café adulterado	5	192	0,9846	0,9669
Predito como sem classe	0	0		
Tabela de confusão ( <b>Predição</b> )				
Classes reais				
	Café puro	Café	Sensibilidade	Especificidade

		Adulterado			
Predito como café puro		18	0	0,6000	1,0000
Predito como café adulterado		12	114	1,0000	0,6000
Predito como sem classe		0	0		

FONTE: O autor (2023).

A Figura 11 mostra a predição feita pelo modelo para cada tipo de conjunto de metodologia usado na validação externa. Considerando apenas os conjuntos de dados na metodologia I, onde foram usadas amostras da mesma origem do que as usadas na calibração, obtemos uma sensibilidade na identificação de café puro e café adulterado de 100%.

Figura 11 - Gráfico de predição de Y predito para grupo Café puro, em destaque as amostras de cada tipo de metodologia de torrefação.



FONTE: O autor (2023).

Em relação a metodologia II, 12 de um total de 18 amostras de café puro foram atribuídas erroneamente ao grupo de café adulterada, isso mostra o impacto da diferença de processos de torrefação na composição química das amostras e consequentemente mudando o espectro NIR.

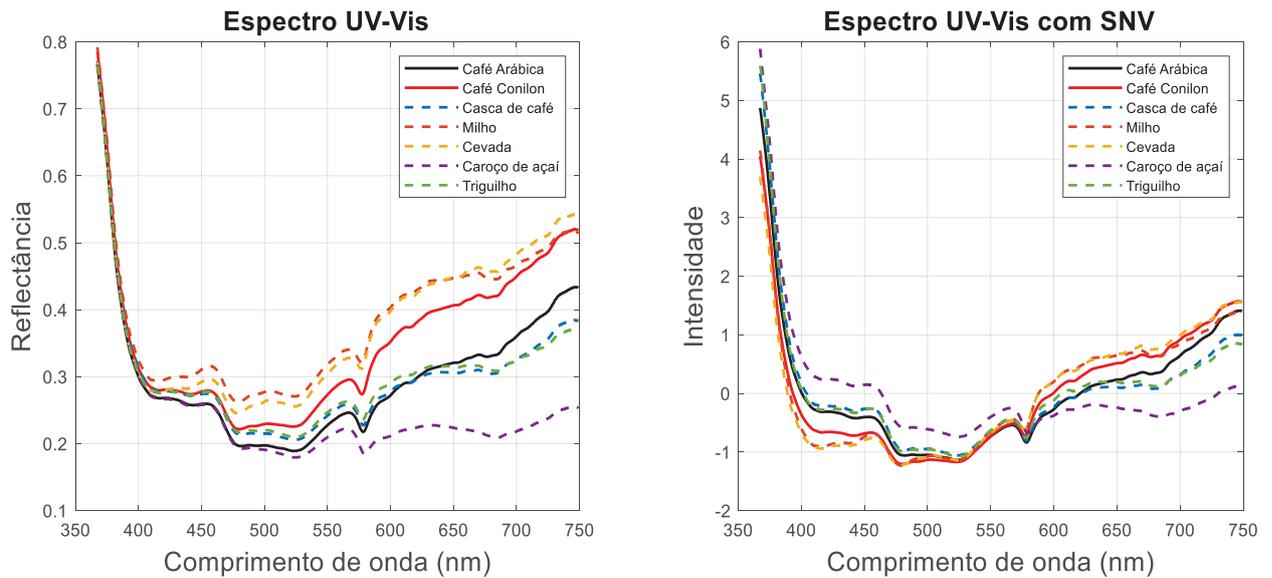
Com isso, o modelo apresentar bons resultados de validação cruzada e até mesmo para validação externa (quando consideramos apenas o conjunto de amostras da metodologia I) com a discriminação de 100 % do café puro para o adulterado. Para fins de comparação, o trabalho de Souto aplica um tratamento da amostra em solução aquosa de café discrimina o grupo puro de adulterado com 94% de sensibilidade usando PLS-DA com um equipamento de espectroscopia UV-Vis.<sup>17</sup>

Todavia, o modelo não performa de maneira eficiente quando amostras produzidas por diferentes metodologias são preditas, implicando que a variabilidade entre métodos tem maior peso neste modelo do que as pequenas concentrações de adulterantes nas amostras. Necessitando de uma maior variabilidade no conjunto de dados para criar um modelo que fosse robusto o suficiente para se adequar em todos os casos de adulteração de café.

## 5.2 ANÁLISE EXPLORATÓRIA DOS DADOS DE HSI-VIS

Para cada uma das 134 amostras, foram coletados um total de 700 espectros em uma área de 5 x 5 mm totalizando um total de 93800 espectros no visível. A Figura 12 apresenta os espectros médios de cada um dos tipos de amostras puras: café arábica, café conilon, casca de café, milho, cevada, caroço de açaí e trigoilho.

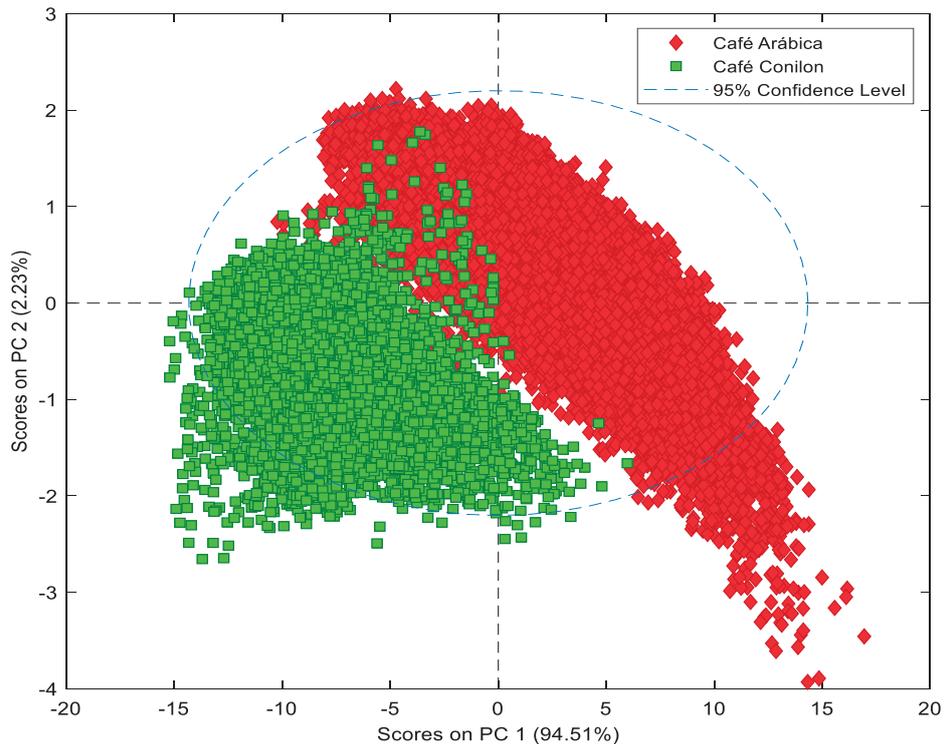
Figura 12 - Espectros UV-Vis médio do café arábica e café conilon em linha sólida, e os adulterantes em linha tracejadas. Do lado esquerdo dados brutos e ao lado direito com tratamento normalização SNV.



FONTE: O autor (2023).

Observa-se grande diferença espectral entre as espécies puras, mesmo estando trabalhando em uma região espectral majoritariamente no visível e com amostras extremamente parecidas ao olho humano. Uma análise exploratória com PCA, utilizando como pré-tratamento SNV e centragem na média, com os espectros de todas as amostras de café arábica (25900 espectros) e café conilon (8400 espectros) obtidas no ensaio inicial, deixa evidente a tendência de separação das espécies. Os resultados de escores na Figura 13 mostram a tendência de separação entre os dois grupos.

Figura 13 - Escores da PCA entre as amostras de café arábica(vermelho) e café conilon(verde).



FONTE: Autoria própria (2023).

A evidente separação entre os tipos de café arábica e conilon é promissor considerando as condições experimentais. Isso é interessante pela aplicação de identificação do tipo de café, sobretudo pelo histórico de adulteração do café arábica por café conilon devido ser um tipo de café mais barato.

### 5.3 MODELOS COM ESPECTROS MÉDIOS

A fim de explorar as diferenças espectrais evidenciadas pela Figura 12 e Figura 13, será tratado cada amostra com a média dos seus 700 espectros adquiridos ponto a ponto na área da amostra.

#### 5.3.1 Modelos PLS-DA com espectros médios

O modelo de discriminação foi construído com a matriz Xcal de calibração com 13440 espectros, 6720 de café arábica puro e 6720 de café conilon puro. Para este modelo, optou-se em trabalhar com uma matriz de validação externa com espectros médios das amostras binárias de café arábica adulteradas (40 espectros)

e amostras de café conilon adulteradas (20 espectros). Os níveis de adulteração são de 3,4,5,10 e 20%, em que o último nível só contém duas amostras. Isso implica que os espectros médios são mais influenciados pela contribuição da componente majoritária da amostra, o café. Além disso, seria possível testar a robustez do modelo para amostras com variações espectrais provenientes de adulterações existentes na amostra, gerando um conjunto de validação significativamente diferente do conjunto de calibração. O pré-tratamento que resultou em uma melhor discriminação, com 2 variáveis latentes, foi SNV e autoescalamento.

A tabela de confusão (Tabela 4), apresentada abaixo, indica que tanto na calibração do modelo como na validação do mesmo foram obtidos resultados de classificação satisfatórios, onde a chance de identificar uma amostra de café arábica como arábica, verdadeiro positivo, e uma amostra de café conilon como conilon, verdadeiro positivo, é de 98,3% e 98,0% (sensibilidade) respectivamente para calibração.

Tabela 4 - Tabela de confusão da calibração do modelo PLS-DA 1.

Tabela de confusão (Validação cruzada)				
	Classes reais		Sensibilidade	Especificidade
	Café arábica	Café conilon		
Predito como café arábica	6606	132	0,9830	0,9803
Predito como café conilon	114	6588	0,9803	0,9830
Predito como sem classe	0	0		
Tabela de confusão (Validação)				
	Classes reais		Sensibilidade	Especificidade
	Café arábica	Café conilon		
Predito como café arábica	39	0	0,9750	1,0000
Predito como café conilon	1	20	1,0000	0,9750

conilon		
Predito como sem	0	0
classe		

Para validação externa, foi utilizado a matriz com espectros médios das amostras adulteradas, qual obteve-se resultados de predição combatíveis com o obtido na calibração, possibilitando distinguir os tipos de café torrados e moídos, com parâmetros de sensibilidade de 97,5% e 100% para café arábica e conilon, respectivamente.

Apesar da dificuldade de encontrar um sistema de análises similar para fazer uma comparação de resultados, o trabalho de Núñez com um equipamento de cromatografia de alta eficiência acoplado a uma UV atingiu valores de discriminação entre as espécies de 91% para amostras de café arábica e café conilon do Vietnã.<sup>58</sup> Logo, além de superior os resultados encontrados neste trabalho, é importante ressaltar a ausência de preparo de amostra.

### 5.3.2 Modelos DD-SIMCA com espectros médios

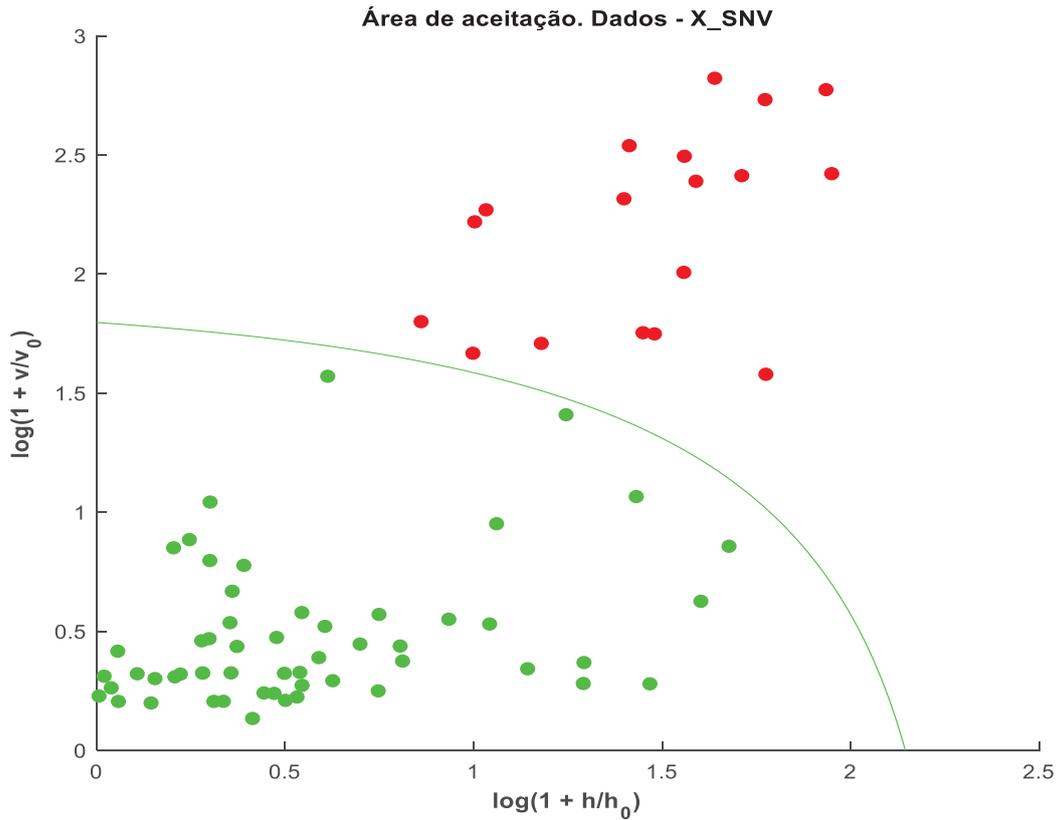
Outra abordagem para identificação das amostras é com modelos de classificação de uma classe (do inglês, one class classification, sigla, OCC), esse por sua vez apresenta uma vantagem frente a modelos como PLS-DA visto necessitar apenas de dados referente ao grupo de interesse para sua discriminação. Desta forma, se futuramente componentes diferentes forem utilizados para adulterar as amostras de café, o modelo por DD-SIMCA continuará preciso para encontrar o grupo de interesse, no nosso caso, café.

Inicialmente, para fins de comparação com o modelo PLS-DA com espectros médios, foi construído um modelo DD-SIMCA, para classificar entre as amostras, qual deles é café arábica. Para calibração foi utilizado os mesmos 6720 espectros de café arábica puro. O pré-tratamento utilizado foi SNV e centragem na média. Duas componentes principais foram usadas para esse modelo e o cálculo da área de aceitação foi feito pela distribuição chi-quadrado. Para validação foi aplicado usado os mesmos espectros médios das 60 amostras.

Na Figura 14, temos a classificação frente ao modelo de classe única. As legendas ao lado dos pontos são as composições reais das amostras. Sendo assim,

percebe-se que nenhuma amostra de café arábica ficou fora da área de aceitação do modelo, todavia duas amostras de café conilon também foram classificadas como café arábica, como é possível verificar pela Figura 14 e Tabela 5.

Figura 14 - Gráfico de aceitação do modelo DD-SIMCA café arábica para as amostras adulteradas. Pontos em verde indicam classificação da amostra como arábica, pontos em vermelho indicam classificação como não arábica.



FONTE : O autor (2023).

Tabela 5 - Resultados modelo DD-SIMCA para café arábica.

Composição real	Área de aceitação		Figuras de mérito	
	Café arábica	Desconhecido	Sensibilidade	Especificidade
Café arábica	40	0	1,0000	0,9524
Café conilon	2	18		

Fonte: O autor (2023).

Resultado similar é obtido fazendo a investigação a partir do café conilon, criando um modelo DD-SIMCA a partir dos espectros puros desse café. Para

calibração foram usados os mesmos 6720 espectros de café conilon puro, os mesmos utilizados na construção do modelo PLS-DA 1, com SNV e centragem na média como pré-tratamentos. O modelo foi construído com 2 componentes principais e a área de aceitação com a distribuição chi-quadrado (Tabela 6).

Tabela 6 - Resultados modelo DD-SIMCA para café conilon.

<b>Composição real</b>	<b>Área de aceitação</b>		<b>Figuras de mérito</b>	
	Desconhecido	Café conilon	Sensibilidade	Especificidade
Café arábica	39	1		
Café conilon	1	19	0,9500	0,9500

FONTE: O autor (2023)

Ambos os modelos apresentaram um valor de sensibilidade acima de 95% para classificar as amostras de café em café arábica e conilon, mesmo na presença de adulterantes, demonstrando a aplicabilidade de utilizar um espectrofotômetro portátil na região do visível para classificar as variedades de café.

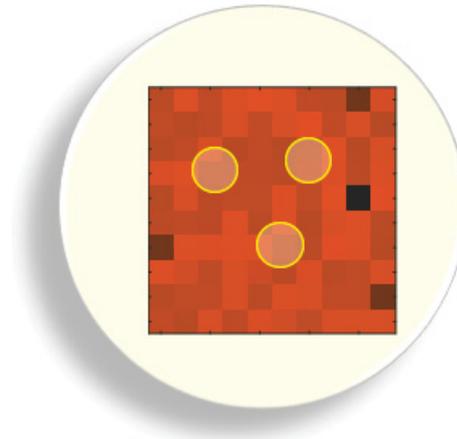
Entretanto, ambos os modelos não foram eficientes na detecção de adulterantes nas amostras (dados não apresentados aqui). Como comentado anteriormente, ao utilizar os espectros médios da imagem, a contribuição espectral dos adulterantes (principalmente os presentes em baixos níveis de adulterações) fica reduzida, quando comparada com a contribuição espectral do café, o que faz com que os modelos de classificação não consigam detectar as variações espectrais referentes aos adulterantes em algumas amostras analisadas.

#### 5.4 IMAGEAMENTO HIPERESPECTRAL UV-VIS

Trabalhar com experimentos de imageamento garante informação relevantes ponto a ponto de determinada imagem. Para o nosso caso, estamos trabalhando com umas misturas heterogêneas, café mais um adulterante, logo realizar apenas uma medição de espectroscópica não garante que vamos pegar a informação média das componentes que constituem nossa amostra. A Figura 15 ilustra o que pode acontecer ao realizar medições espectroscópicas em uma amostra. Os círculos em

amarelos são referentes a posição de cada medição, nota se que em nenhuma das réplicas foi possível encontrar a informação referente ao ponto mais escuro, com isso o analista pode ter uma falsa impressão de uma amostra pura, onde na realidade a aleatoriedade das medições ocasionou resultados homogêneos.

Figura 15 - Simulação de aquisição de espectros em 3 posições diferentes da amostra (círculos amarelos). A diferença na colocação dos pixels indicam composições.



FONTE: O autor (2024).

Para contornar esse problema, a técnica de imageamento traz a possibilidade de mapear uma área da amostra e correlacionar a posição no espaço a uma informação química relevante, assegurando apontar a localização de informações químicas diferentes.

#### 5.4.1 Modelos DD-SIMCA

Buscando responder se existe adulteração na amostra de café e localizar espacialmente essa adulteração, não importando o tipo de adulterante, foi elaborado modelos DD-SIMCA para cada espécie. Os modelos foram treinados com espectros dos imageamentos das suas respectivas amostras puras. Será apresentada aqui apenas os resultados com o melhor pré processamento, snv e autoescalamento, e com número otimizado de componentes principais.

Para definir o número ideal de componentes, foi feito modelos entre 1 a 10 componentes principais e calculado a sensibilidade de cada um deles, escolhendo

no final o como maior valor. A Tabela 7, apresenta os resultados dos modelos que serão usados nas amostras adulteradas.

Tabela 7 - Número de espectros usados na construção e validação de cada um dos modelos DD-SIMCA.

Espécie	Calibração (espectros)	Validação (espectros)	Número de componentes	Sensibilidade	Especificidade
Café arábica	5376	1344	3	0,9807	0,9897
Café conilon	5376	1344	3	0,974	0,9349
Casca de café	1680	420	5	0,9929	0,8721
Milho1	560	140	1	0,9928	0,9917
Milho2	1120	280	3	0,9929	0,9726
Cevada	560	140	1	0,9929	0,9976
Caroço de açaí	1680	420	4	0,9762	0,9995
Triguilho1	560	140	4	0,9857	1
Triguilho2	560	140	1	0,9786	0,9782

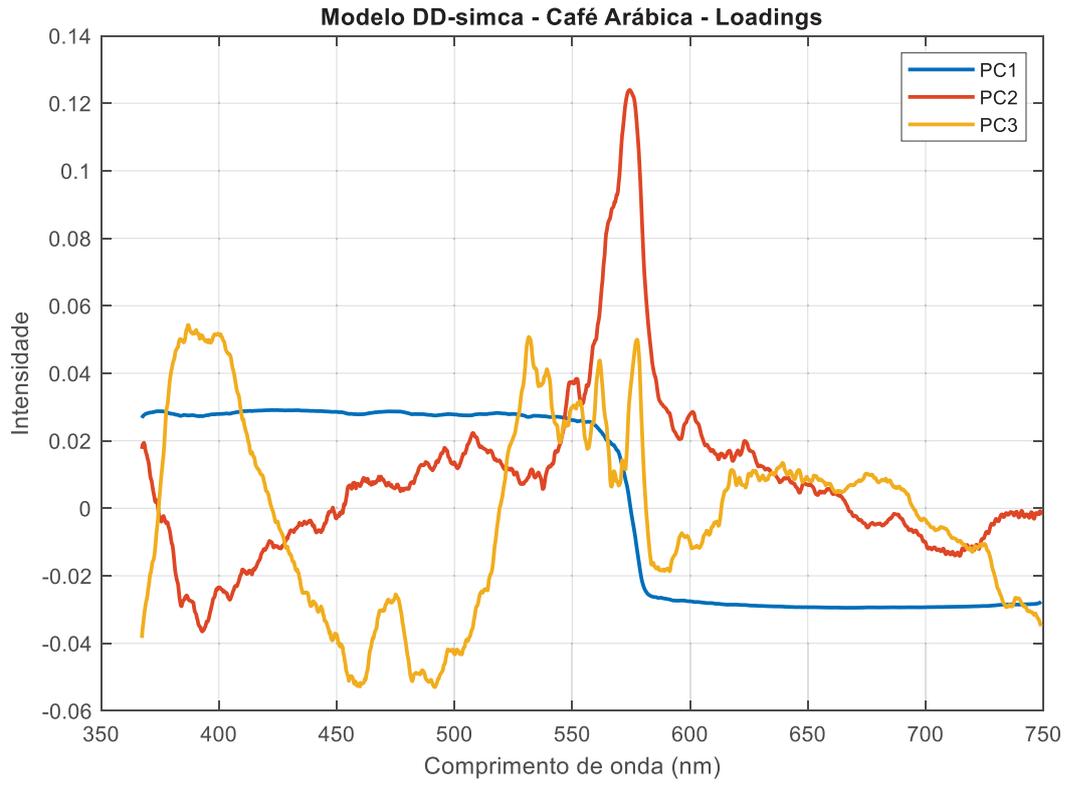
FONTE: O autor (2024).

A escolha do número de componentes foi em função da melhor performance dos modelos avaliadas com base nas figuras de mérito sensibilidade e especificidade. Para todas as espécies foi possível obter modelos com figuras de mérito satisfatórias.

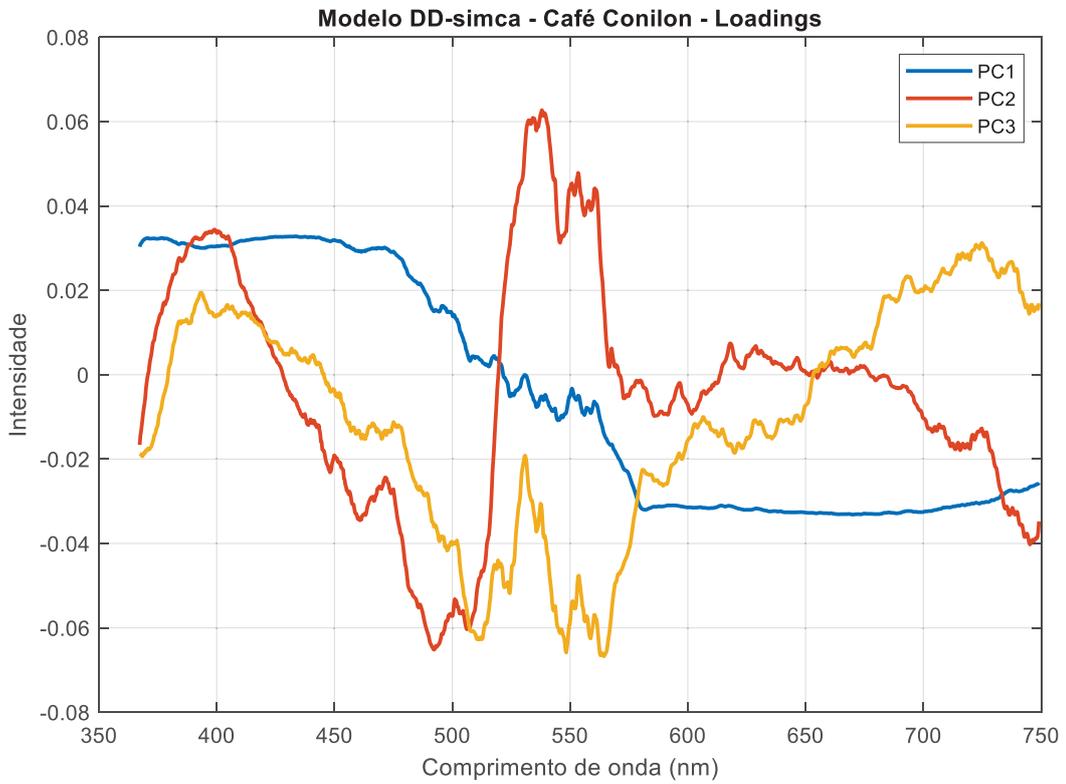
Comparando os *loadings* das componentes principais do modelos de café arábica e café conilon notasse uma diferença significativa entre os tipos de café (Figura 16).

Figura 16 - Loadings dos modelos DD-SIMCA para a)café arábica e b)café conilon.

a)



b)



FONTE: O autor (2024).

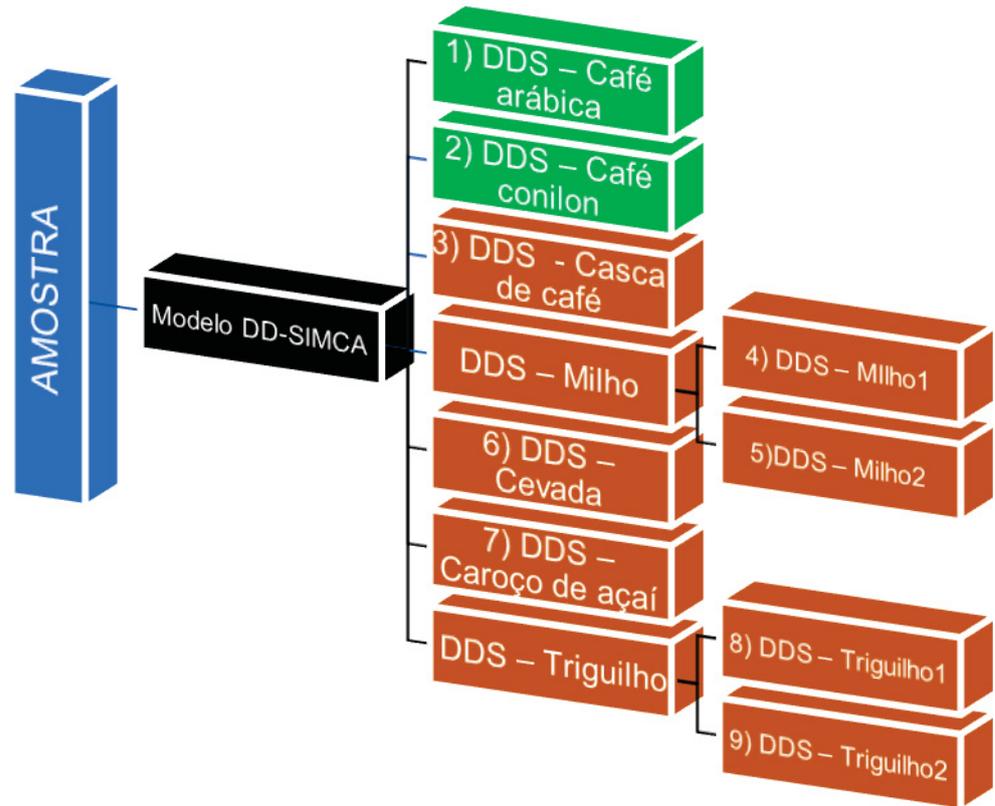
A semelhança entre as componentes principais 1 de ambos os modelos indicam que essa variabilidade está associada a uma variância do equipamento. Outro ponto de atenção é a banda em 575 nm associado a segunda componente principal para o modelo de café arábica, esse é o mesmo comprimento de onda em que temos a mudança de sinal na componente principal 1, qual foi atribuída a variância do equipamento, indicando então possivelmente o pronunciamento da influência do equipamento na segunda componente para café arábica.

Comparando as componentes principais entre os modelos, observamos grande diferença entre 400 a 750 nm, principalmente nas componentes 2 e 3. O café conilon apresenta diferenças mais significativas e intensas ao longo da faixa espectral.

#### 5.4.1.1 Aplicando os modelos DD-SIMCA nas amostras adulteradas.

O processo de investigação da composição da amostra foi de acordo com o fluxograma presente na Figura 17. O imageamento de cada uma das amostras passou pela predição dos 9 modelos, dois para os tipos de café, e 7 para os possíveis adulterantes.

Figura 17 - Fluxograma modelos DD-SIMCA.



FONTE: O autor (2024).

Seguindo essa lógica, para uma amostra de café arábica, por exemplo, esperava-se grande quantidade de pixels (espectros) classificados como café arábica, e um número pequeno de pontos desconhecidos. Esses, por sua vez poderiam estar classificados na área de aceitação de modelos referente aos adulterantes.

De forma a demonstrar os resultados obtidos nas 73 amostras, a amostra id\_9732422, composta de café Arábica e milho será usada como exemplo para descrição do processo e para mostrar o potencial da ferramenta.

Com os 700 espectros de imageamento da amostra, foi predito para cada modelo DD-simca. A tabela abaixo mostra quantos espectros foram preditos como dentro da variabilidade mapeada por determinado modelo. Nesse caso será tratado como dentro, ou fora do espaço de aceitação do modelo respectivamente.

Tabela 8 - Resultados referente a amostra de café arábica e milho (id : 9732422) ao ser aplicado nos modelos DD-simca.

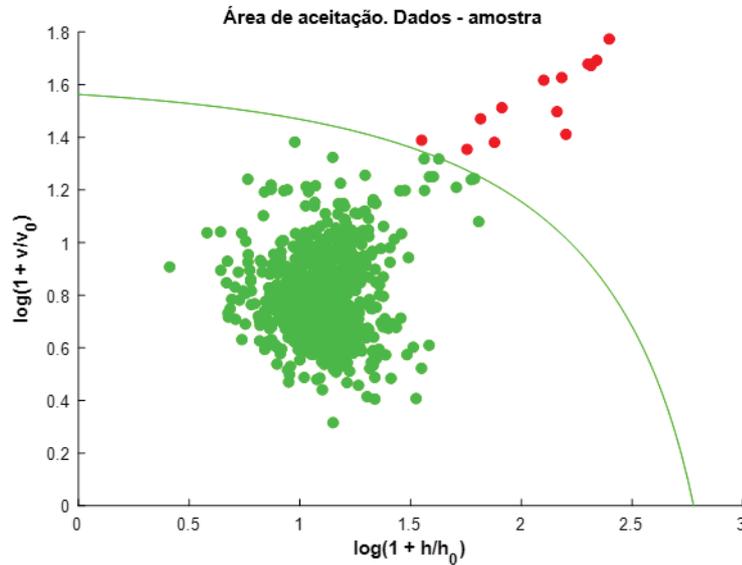
Modelos	Espectros	Amostra id_9732422
DDS-Arábica	Dentro	682

	Fora	18
<b>DDS-Conilon</b>	Dentro	0
	Fora	700
<b>DDS-Casca</b>	Dentro	3
	Fora	697
<b>DDS-Milho1</b>	Dentro	0
	Fora	700
<b>DDS-Milho2</b>	Dentro	0
	Fora	700
<b>DDS-Cevada</b>	Dentro	0
	Fora	700
<b>DDS-Caroço</b>	Dentro	0
	Fora	700
<b>DDS-Triguiho1</b>	Dentro	0
	Fora	700
<b>DDS-Triguiho2</b>	Dentro	0
	Fora	700

FONTE: O autor (2024).

Convergindo com a informação da composição real, 687 dos 700 espectros da amostra estão dentro da área de aceitação do café arábica. Isso nos indica que a amostra é composta majoritariamente por esse tipo de café, e os demais espectros, 13, são diferentes de café arábica, pouco importando o que é. Na Figura 18 O gráfico de aceitação para essa amostra é apresentado.

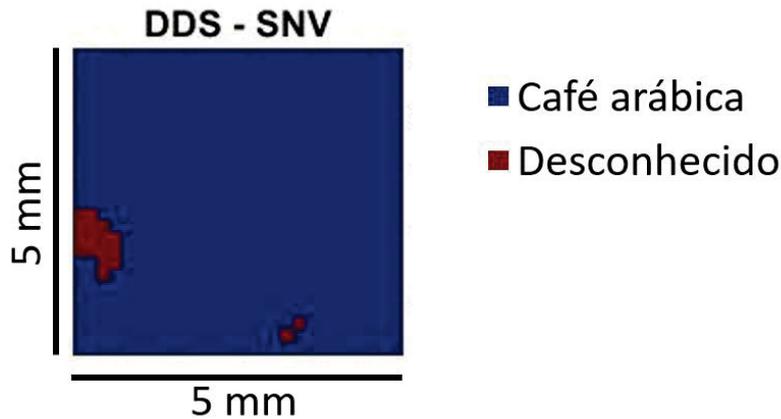
Figura 18 - Área de aceite calculada pelo modelo DDS-café arábica. Os pontos verdes indicam espectros semelhança com modelo, pontos vermelhos indicam espectros diferente de café arábica.



FONTE: O autor (2024).

Uma melhor maneira para visualizar esse resultado é remontando os 700 espectros como um cubo, recuperando assim a informação espacial associada a cada espectro. Com a Figura 19 podemos observar onde está localizado os pontos da amostra que foram classificados fora do modelo de café arábica.

Figura 19 - Imageamento da amostra com predição do modelo DD-SIMCA café arábica.



Fonte: O Autor (2024).

Associando o resultado do modelo DD-simca com a posição espacial de cada espectro temos um forte indício de que os espectros que não foram classificados como café arábica, podem de fato ser uma espécie estranha ao café arábica, um possível adulterante. Em seguida, seguindo o fluxo descrito na Figura 17, buscando identificar os espectros rejeitados pelo modelo de café arábica, os

dados da amostra foram preditos com os modelos dos adulterantes buscando encontrar uma área complementar ao da figura acima, preenchendo os pontos vermelhos encontrados pelo modelo de café arábica.

Entretanto, apenas o modelo de casca de café pontuou 3 espectros como sendo casca de café, sem coerência com a composição da amostra que é café arábica e milho. É interessante apontar que para quase todas as amostras, não foi encontrado pontos associados aos modelos de milho, cevada, caroço e trigoilho, mesmo para amostras contendo tais adulterantes.

Uma vez feito isso para todas as amostras com adulterantes, foi investigado cada uma das suas áreas mapeadas. Pela Figura 20, das 40 amostras de café arábica adulteradas com apenas um tipo de adulterante, 29 amostras foram indicadas como puras. Por outro lado, para as 20 misturas de café conilon com apenas um adulterante, 5 foram indicadas como puras.

Figura 20 - Quantidade de amostras indicadas como puras e adulteradas por seus respectivos modelos DD-SIMCA.

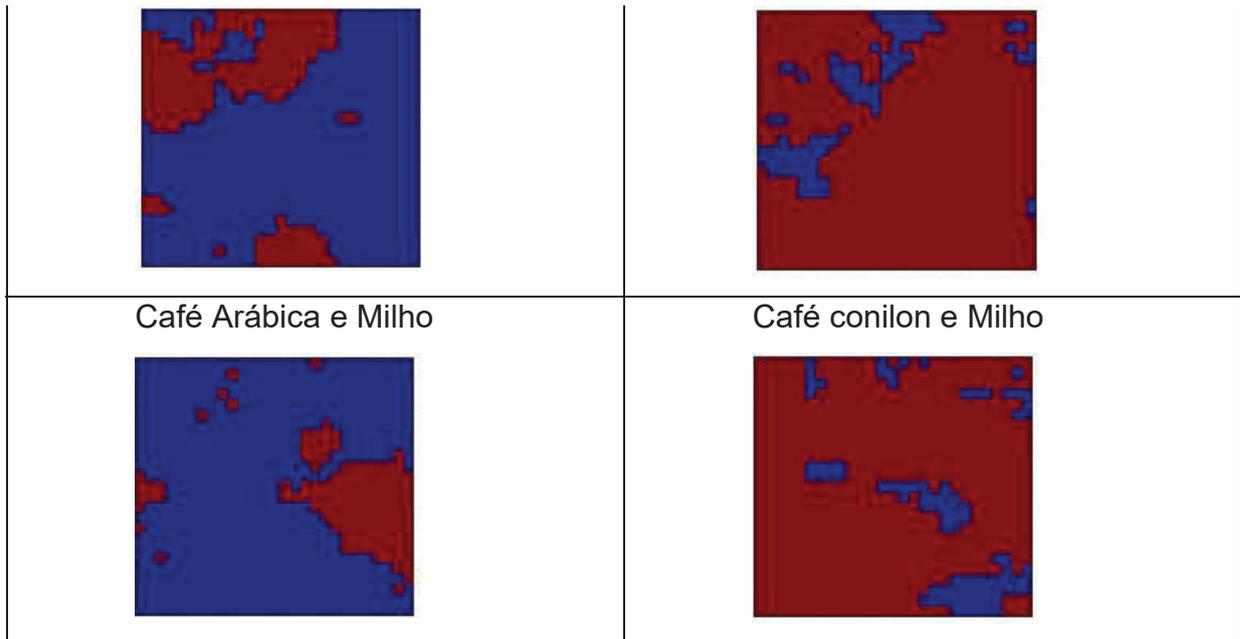


FONTE: O autor (2024).

Abaixo a Figura 21 é apresentado outros exemplos da predição dos mapas para cada tipo de café das amostras

Figura 21 - Mapas preditos pelos respectivos modelos de café das amostras. Em azul são áreas atribuídas ao tipo de café, e em vermelho pontos desconhecidos.

DD-SIMCA	
Café Arábica e Casca de café	Café Conilon e caroço de açaí



Apesar do modelo de café arábica ter sensibilidade maior do que do café conilon, esse recuperou mais mapas das amostras adulteradas confirmando possíveis espectros suspeitos. Isso pode ser reflexo de atribuição da especificidade do modelo de café conilon de 93%, podendo atribuir mais espécies desconhecidas ao modelo com o café, gerando um resultado falso positivo.

Isso pode impactar no número de total de espectros atribuídos como desconhecidos, comprometendo uma análise quantitativa, todavia para análises qualitativas, o modelo de café conilon apresenta mais coerência com a informação real do número de amostras adulteradas.

#### 5.4.2 Modelos PLS-DA

Outro método supervisionado para identificação e distinção entre grupos é o PLS-DA. Diferentemente do DD-SIMCA, precisamos ensinar para o modelo quais grupos queremos discriminar, logo o conjunto de calibração e validação precisar contar a variabilidade de cada um dos grupos.

Dois modelos serão apresentados, um para discriminar café arábica puro de adulterantes, sendo estes café conilon, casca de café, milho, cevada, caroço de açaí e trigoilho. O segundo modelo, vamos diferenciar café conilon e adulterantes compostos café arábica, casca de café, milho, cevada, caroço de açaí e trigoilho. Vale ressaltar que o café arábica não é adulterante, vamos manter ele no grupo dos

adulterantes nesse caso apenas para ensinar que a variabilidade do café arábica não é desejada no grupo de café conilon.

Na Tabela 9 temos a configuração de cada um dos modelos, indicando o número de espectros usados em cada conjunto de calibração e validação.

Tabela 9 - Configuração do conjunto de dados usados para calibração e validação dos modelos.

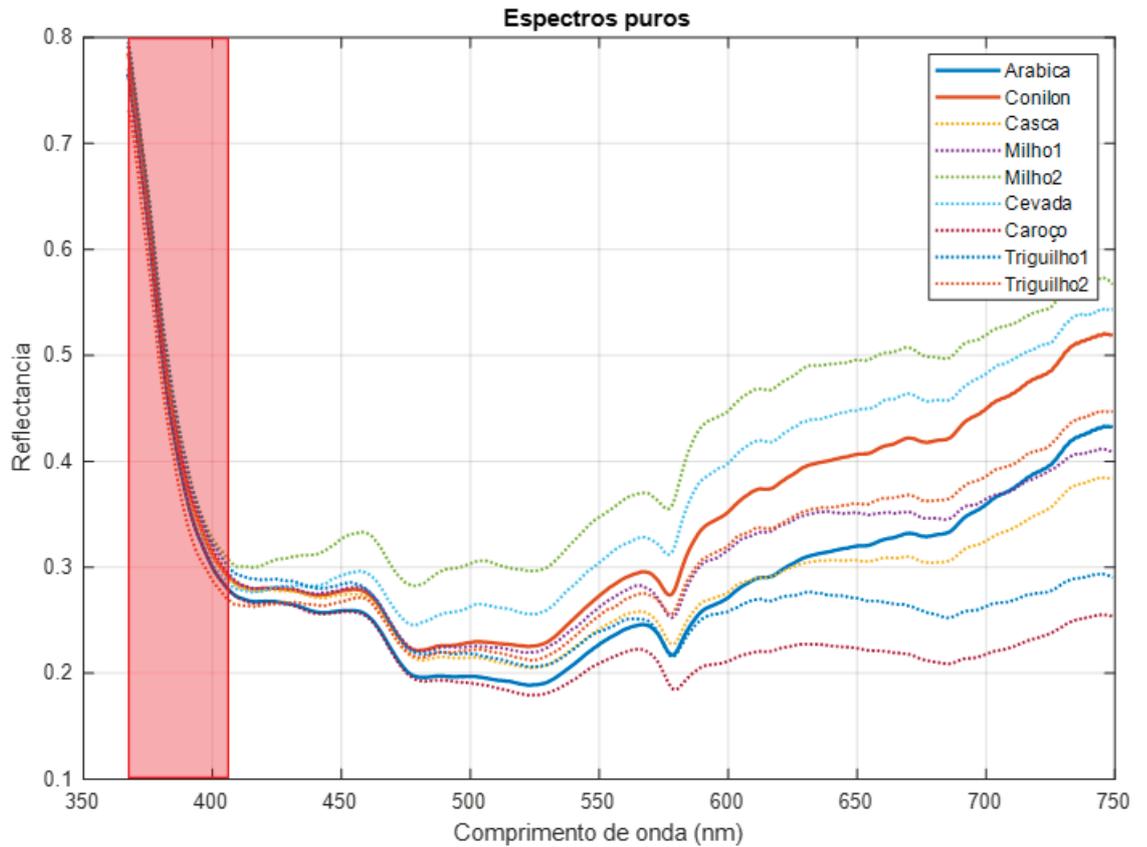
s	Modelo Café arábica		Modelo Café conilon	
	Café Arábica	Adulter antes <sup>1</sup>	Café Conilon	Adulter antes <sup>2</sup>
Calibra	5376	4480	5376	4480
ção	1344	1120	1344	1120
<p>Adulterantes<sup>1</sup>: Café conilon, casca de café, milho1, milho2, cevada, caroço de açaí, trigoilho1 e trigoilho2.</p> <p>Adulterantes<sup>2</sup>: Café arábica, casca de café, milho1, milho2, cevada, caroço de açaí, trigoilho1 e trigoilho2.</p>				

FONTE: O autor (2024).

A quantidade usada de adulterantes para os conjuntos de calibração foi de 560 espectros de cada espécie, buscando ter um equilíbrio da quantidade de informação do grupo de café e do adulterante. O conjunto de validação também foi montado buscando manter as proporções entre os dois grupos de cada modelo para evitar tendências.

Uma mudança feita em relação aos modelos anteriores foi um corte das primeiras 176 variáveis. A Figura 22 abaixo mostra a região que não foi usada para criação dos modelos PLS-DA.

Figura 22 – Corte espectral representado pela área vermelha.



FONTE: O autor (2024).

Essa alteração foi feita para evitar maior peso para essas variáveis iniciais visto que a sua variabilidade é mais sensível ao equipamento do que a diferença espectral entre as amostras. Com isso, será usado os canais espectrais entre 419 a 719nm.

Em relação ao pré-processamento, será apresentado somente o modelo com normalização por SNV e autoescalonamento visto que não houve melhoras significativas nas figuras de mérito. As tabelas abaixo trazem as figuras de mérito para os dois modelos elaborados, Tabela 10 e Tabela 11 para modelo de café arábica e adulterantes e café conilon e adulterantes respectivamente

Tabela 10 - Figuras de mérito para modelos PLS-DA para café arábica.

Modelo- Café arábica – 5 Variáveis latentes				
	Validação cruzada		Predição	
	Sensibilidade	Especificidade	Sensibilidade	Especificidade

Café arábica	2	0	81	0.992	0.9929
Adulter ante	0	2	29	0.99	0.9881

FONTE: O autor (2024).

Tabela 11 - Figuras de mérito para modelos PLS-DA para café conilon.

Modelo- Café conilon – 7 Variáveis latentes				
	Validação cruzada		Predição	
	Sensibilidade	Especificidade	Sensibilidade	Especificidade
Café conilon	0.9665	0.9199	0.9695	0.9321
Adulter ante	0.9199	0.9665	0.9321	0.9695

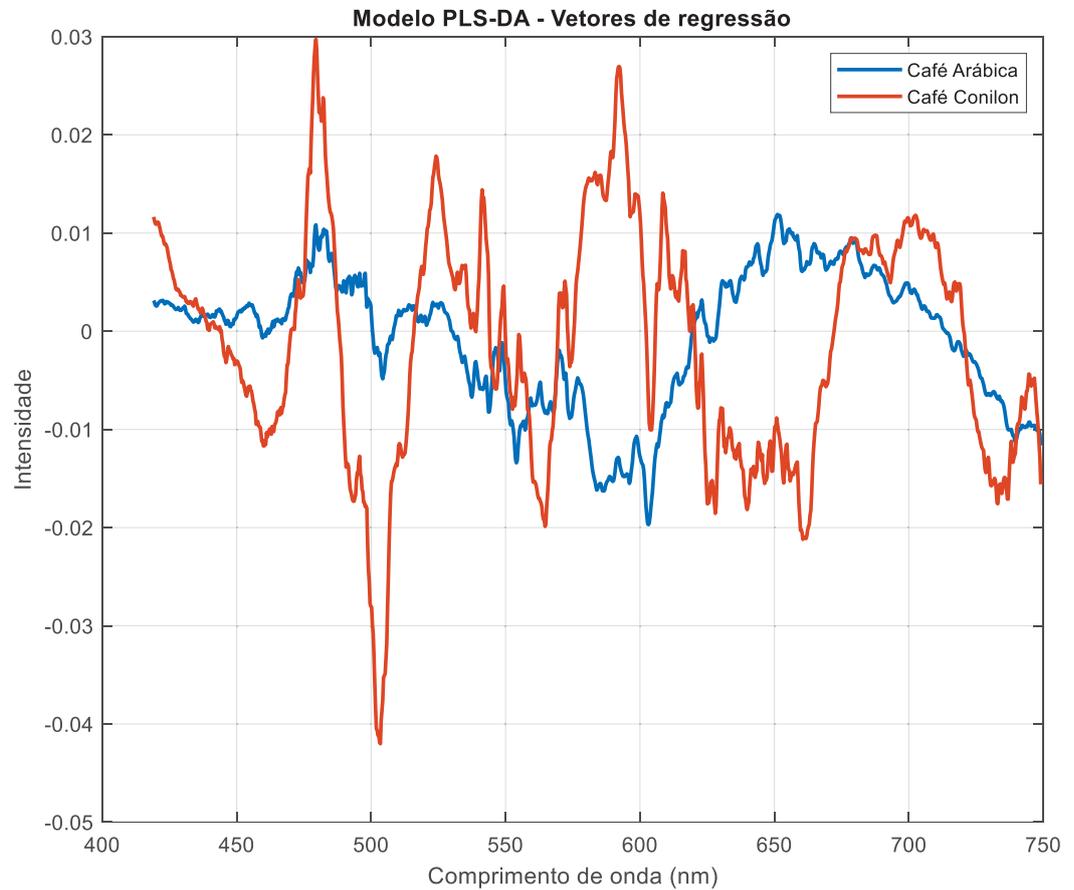
FONTE: O autor (2024).

As figuras de mérito de cada um dos modelos são satisfatórias. Para o modelo de café arábica, com sensibilidade de 98,81% para o conjunto de dados de predição, o café arábica foi classificado como café arábica 98,81% das vezes, casos de verdadeiro positivo. Outro modo de avaliar os modelos é considerando a especificidade, em uma situação real o caso de atribuir um adulterante como café é extremamente problemático, e para entender essa possibilidade observamos a figura de mérito especificidade, onde para o café arábica foi de 99,29%, isso é, o adulterante foi classificado como tal 99,29% das vezes, dando segurança na resposta, evitando casos de adulterantes sendo classificado como café e emissões de laudos com falsos negativo.

Para os modelos de café conilon, obteve-se figuras de mérito inferiores, mas ainda satisfatórias, como 96,95% e 93,21% de sensibilidade para o conjunto de predição dos grupos café conilon e adulterantes, respectivamente. Para esse caso, existe maior probabilidade de um adulterante ser classificado como café conilon, visto a especificidade para o café conilon ser de 93,21%.

Comparando os vetores de regressão dos dois modelos na Figura 23, notasse uma diferença significativa entre os vetores dos modelos, coerente com o que já foi observado nas análises exploratórias como PCA.

Figura 23 - Vetor de regressão dos modelos PLS-DA. Em azul para café arábica e em laranja para café conilon.



FONTE: O autor (2024).

Outro ponto que chama atenção é a diferença na intensidade entre os cafés. Para conilon observa pontos mais definidos e mais intenso enquanto que para o café arábica há menor diferença entre os extremos.

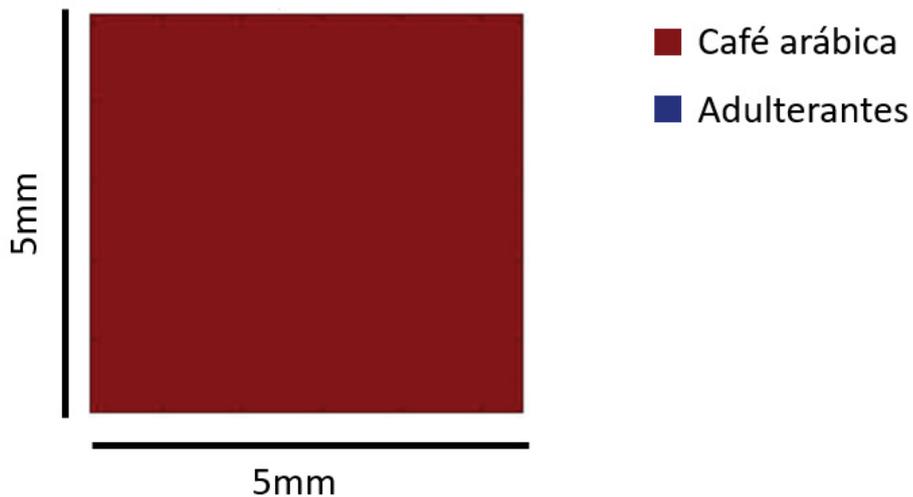
#### 5.4.2.1 Aplicando os modelos PLS-DA nas amostras adulteradas

Com os modelos construídos, cada amostra adulterada foi predita segundo cada uma desses buscando encontrar quantos espectros eram atribuídos a cada grupo e posteriormente associar cada espectro a sua posição no espaço.

Para manter um padrão, vamos usar a mesma amostra contendo café arábica e milho usada para estudo de caso com os modelos DD-SIMCA (Figura 19).

Pela Figura 24, observamos que foi somente recuperado pontos associados ao café arábica, não identificando nenhum adulterante.

Figura 24 - Mapa de predição da amostra de café arábica e milho pelo modelos PLS-DA de café arábica.

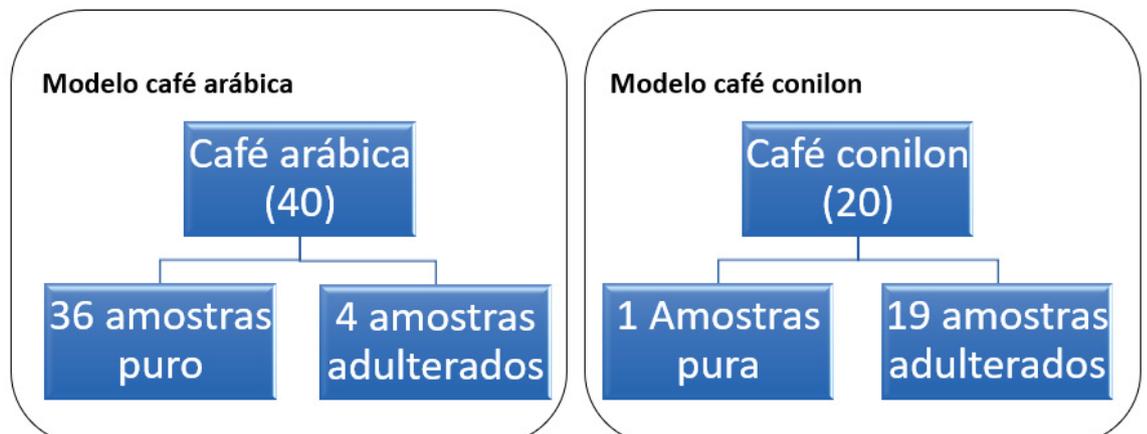


FONTE: O autor(2024).

A identificação de todos os espectros com café arábica é coerente com a composição real da amostra, todavia não encontrar nenhum adulterante entra em divergência com a informação dada pelo modelo DD-simca e pela informação da composição real, que continha adulteração.

O desempenho geral dos modelos para as amostras adulteradas está apresentado na Figura 25 .

Figura 25 - Desempenho da classificação das amostras por seus respectivos modelos PLS-DA.

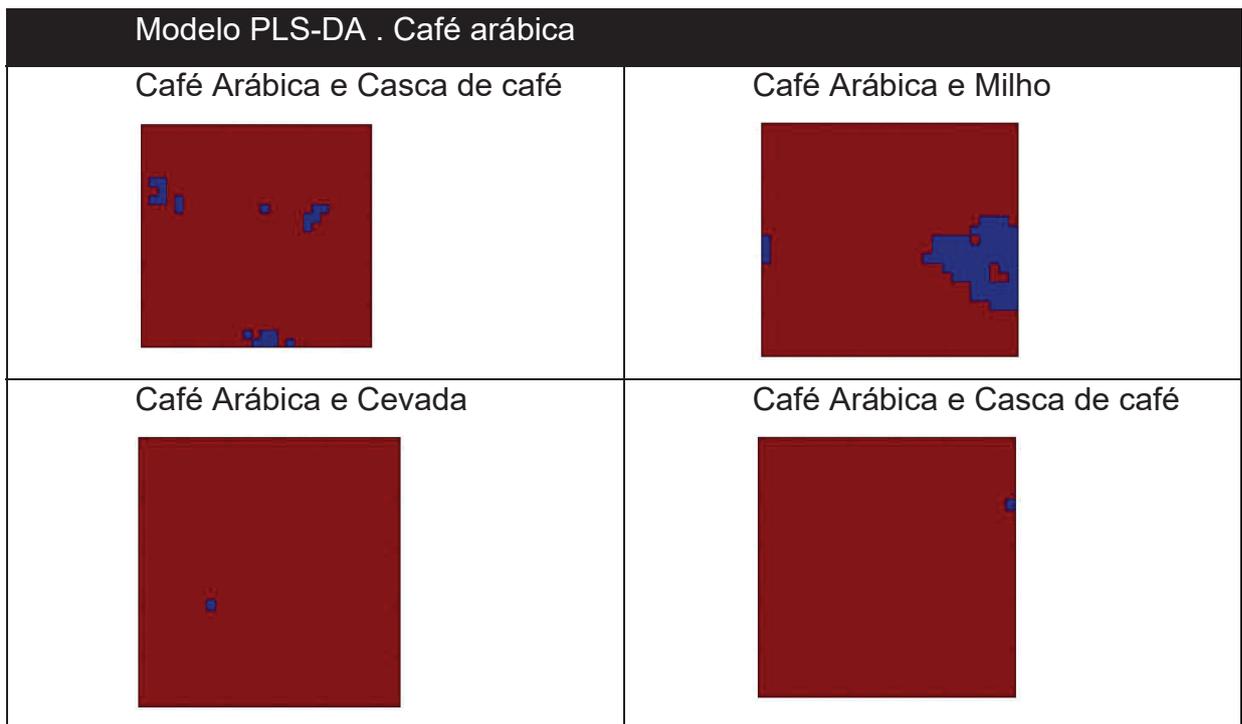


FONTE: O autor (2024).

Analisando individualmente, para café arábica observamos uma piora em relação ao modelo criado com DD-simca, abaixo observamos os 4 mapas que encontrou possíveis adulterantes.

A Figura 26 apresenta os 4 mapas encontrados pelas predições.

Figura 26 - Mapas de predições PLS-DA para café arábica.



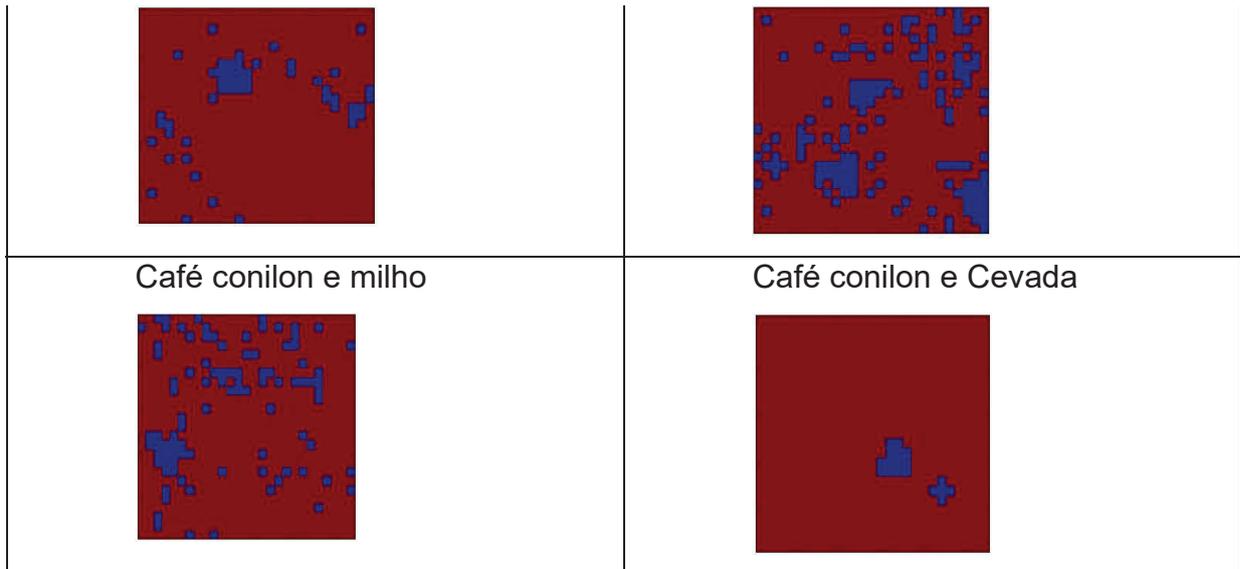
FONTE: O autor (2024).

Os locais onde foi encontrado adulterantes é um indicativo de uma área suspeita, um possível aglomerado de matéria estranha ao café. A quantidade de pixels atribuído aos adulterantes é fortemente influenciada pela resolução espacial de cada aquisição de espectros, isso é, dificilmente um espectro vai ser apenas de uma espécie pura, seja ela adulterante ou café.

Em relação aos modelos de café conilon, das 20 amostras adulteradas foi possível encontrar adulterantes em 19. A Figura 27 seleciona alguns mapas das imagens.

Figura 27 - Mapas de predições PLS-DA para café conilon.





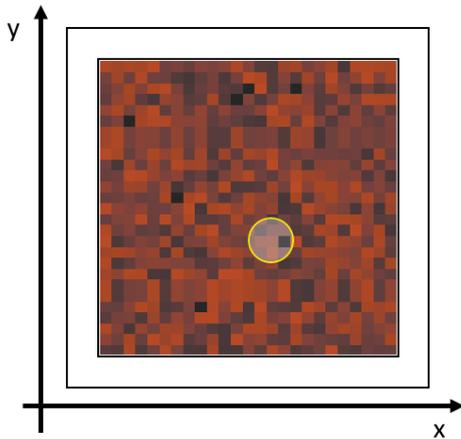
FONTE: O autor (2024).

Para as demais amostras de café conilon, os mapas preditos ficaram mais dispersos, indicando maior distribuição dos adulterantes.

Comparando o resultado dos mapas recuperados com DD-SIMCA e PLS-DA, observamos um comportamento parecido, onde as figuras de mérito indicam um ótimo desempenho do café arábica frente ao modelo de café conilon, todavia ao predizer o mapeamento das amostras adulteradas observamos pouca recuperação da área adulterada em comparação entre café arábica e conilon. Isso é coerente quando observamos a especificidade dos modelos para os seus tipos de café, a chance do modelo de conilon gerar resultados falso positivos é maior, conseqüentemente, o mapa recuperado por ter espectros atribuídos como adulterante, onde não são realmente. Isso compromete diretamente uma aplicação quantitativa, todavia, para análise qualitativa, os mapas ao localizar possíveis adulterantes já é um avanço em relação aos métodos hoje utilizados para fazer a mesma função.

Outro ponto que explica os resultados de recuperação dos mapas das amostras adulteradas é a limitações instrumentais. A Figura 28 simula uma aquisição de uma amostra de café fictícia, sendo que os pontos em marrom representam um tipo ordinário de café e os pontos mais escuros ao longo da área representam adulterantes, o círculo amarelo na figura é a área de aquisição do espectro.

Figura 28 - Imagem simulada de uma amostra de café. Em marrom, temos café e os pontos mais escuros, tendendo a preto, são de adulterantes. O círculo amarelo no meio é área de coleta dos espectros.



FONTE: O autor (2024).

Nesta área, temos a presença majoritária de café e apenas um pixel referente ao adulterante. Isso irá gerar um espectro final que é uma média ponderada das componentes que compõe esta área de aquisição, logo, visto que a presença de adulterantes é baixa, obtemos um espectro final mais parecido com os espectros do café majoritário. Apesar da alta resolução do equipamento,  $200\mu\text{m}$  por pixel, não há como garantir que um espectro seja 100% composto de apenas uma componente, sobretudo com adulterantes variando entre 3 a 20% (m/m).

Desta forma, ao prever uma amostra usando modelos PLS-DA, que foi construído com espectros puros de café e de adulterantes, o espectro da amostra terá maior semelhança ao tipo de espécie majoritária na área em que foi adquirido o espectro, e considerando a baixa concentração dos adulterantes, o resultado final é maior contribuição do tipo de café.

Isso fica mais em evidência para modelos PLS-DA, onde foi mapeado como é a variabilidade do café e do adulterante. No caso do DD-SIMCA, ao indicar apenas a contribuição do tipo de café, pouco importa como é um adulterante, logo, é mais fácil de um espectro da amostra com contribuição de um adulterante ser diferente o suficiente do espectro puro do café e ser classificado como fora do grupo de café.

#### 5.4.3 MCR-ALS

Até o momento, todas as ferramentas quimiométricas usadas foram métodos supervisionados, isso é, necessita ter informação do(s) grupo(s) de interesse. O desempenho do modelo vai ser sensível a informação inicial, logo, se existir diferenças entre o que foi usado na calibração com o que deseja ser previsto, o resultado não será satisfatório.

No nosso caso, observamos até aqui um ótimo desempenho dos modelos para amostras puras, tanto para o tipo de café quanto para os adulterantes. Todavia, ao ser aplica-los para prever as amostras adulteradas observamos dificuldade em localizar os adulterantes devido à natureza baixa concentração dos adulterantes e a dificuldade em produzir uma mistura homogênea entre os compostos. Logo, os espectros das amostras adulteradas têm influência do tipo de adulterante, todavia sua baixa concentração faz com que variância de um espectro puro de café com um espectro da mistura café mais adulterante é menor do que a diferença entre essas espécies puras, logo, prejudicando modelos supervisionados como PLS-DA e DD-SIMCA.

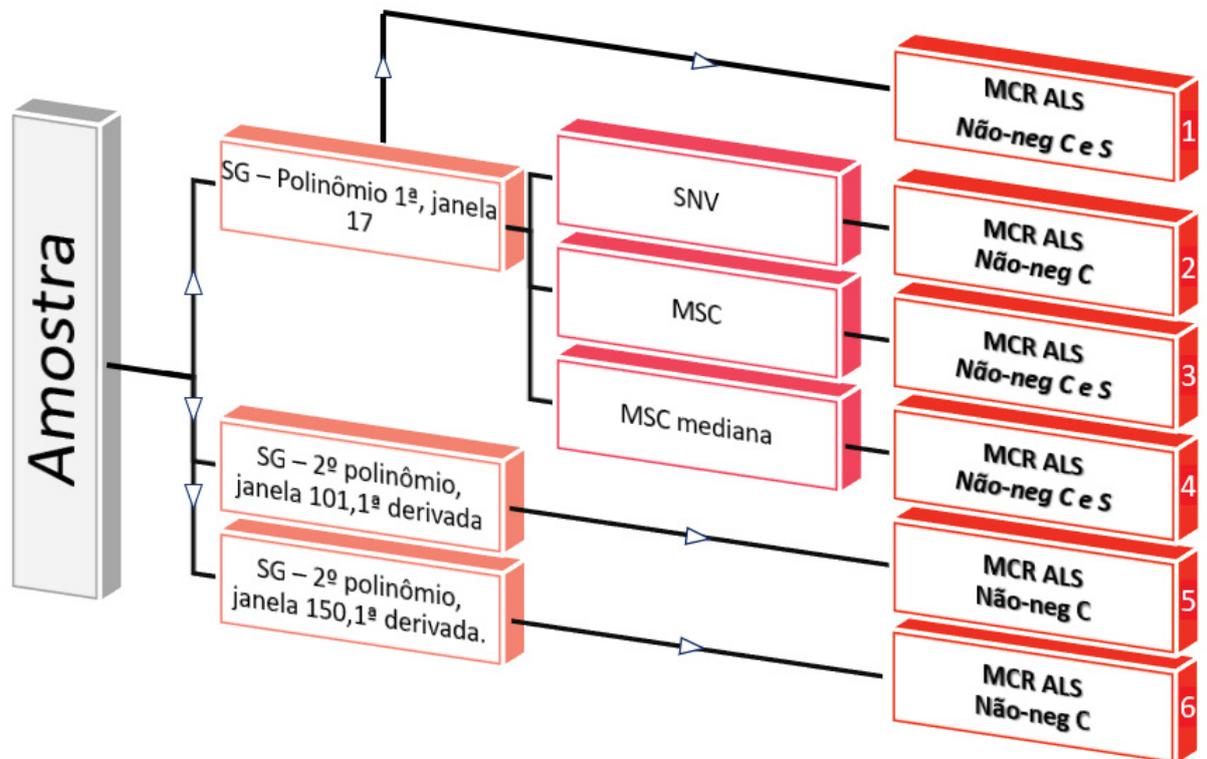
Portanto, para estudar a variabilidade dentro da amostra adulterada, vamos usar a ferramenta MCR-ALS, um método não-supervisionado, onde irá mapear a amostra com base na distribuição da concentração das suas componentes puras.

Um dos destaques dessa ferramenta é a sua resposta com sentido químico. Isso é, ao passar um conjunto de espectros de uma mistura em diferentes concentrações, o MCR-ALS é capaz de encontrar o perfil espectral puro de cada uma das espécies que compõem a mistura. E, quando aplicado em dados de imageamento espectral, o perfil de concentração de cada uma dessas componentes puras está vinculado a um ponto no espaço, sendo possível localizar a distribuição das espécies.

Para iniciar os cálculos pela ferramenta, a estimativa inicial foi definida usando o algoritmo SIMPLISMA, buscando encontrar os espectros mais distintos dentro dos 700 espectros de cada amostra. O número de componentes iniciais foi estipulado pelo conhecimento da quantidade real de componentes puras que constituem a amostra, por exemplo, para uma mistura ternária, foram calculados três componentes iniciais, para uma mistura binária, duas componentes.

Foram executados em total 6 MCR-ALS, cada um com um tratamento diferente, como esquematizado na Figura 29.

Figura 29 - Configuração dos MCR-ALS executados e testes de similaridade com o perfil espectral recuperado.



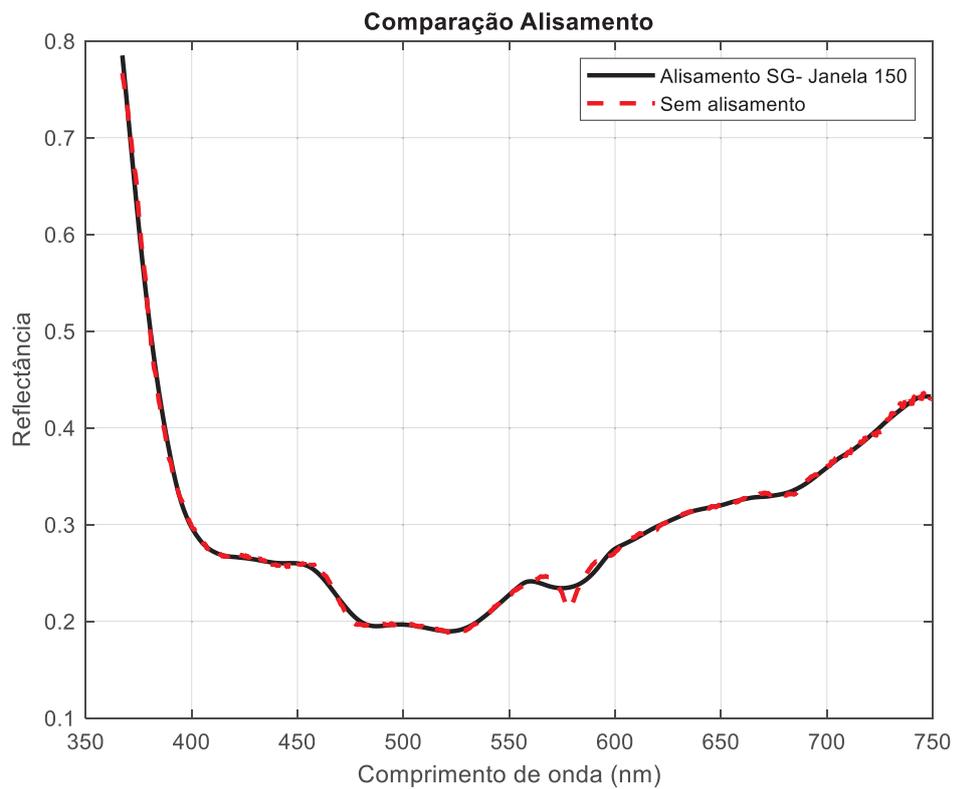
FONTE: O autor (2024).

Para facilitar o entendimento de cada um dos MCR-ALS feitos, vamos chamar cada um deles pelo dígito no final do fluxo da figura acima.

Os resultados de MCR-ALS 1,2,3 e 4, todos passaram pelo mesmo processamento inicial, alisamento dos espectros por Savitzky-Golay (SG) com uma janela de 17 variáveis e polinômio de 1º grau. Para o modelo 2,3 e 4 foi adicionado um processamento de normalização dos dados há mais como indicado no fluxograma. Por fim, para os MCR-ALS 5 e 6 foi aplicado um alisamento de Savitzky-Golay (SG) com uma janela de 101 e 150, respectivamente, polinômio de 2º grau e 1ª derivada.

Vale a pena ressaltar que, apesar da grande janela de alisamento para os MCR-ALS 5 e 6, o espectro sem nenhum tratamento é bastante ruidoso e tem resolução de 0,29nm por canal espectral (Figura 30)

Figura 30 - Comparação entre espectros sem alisamento e com alisamento.



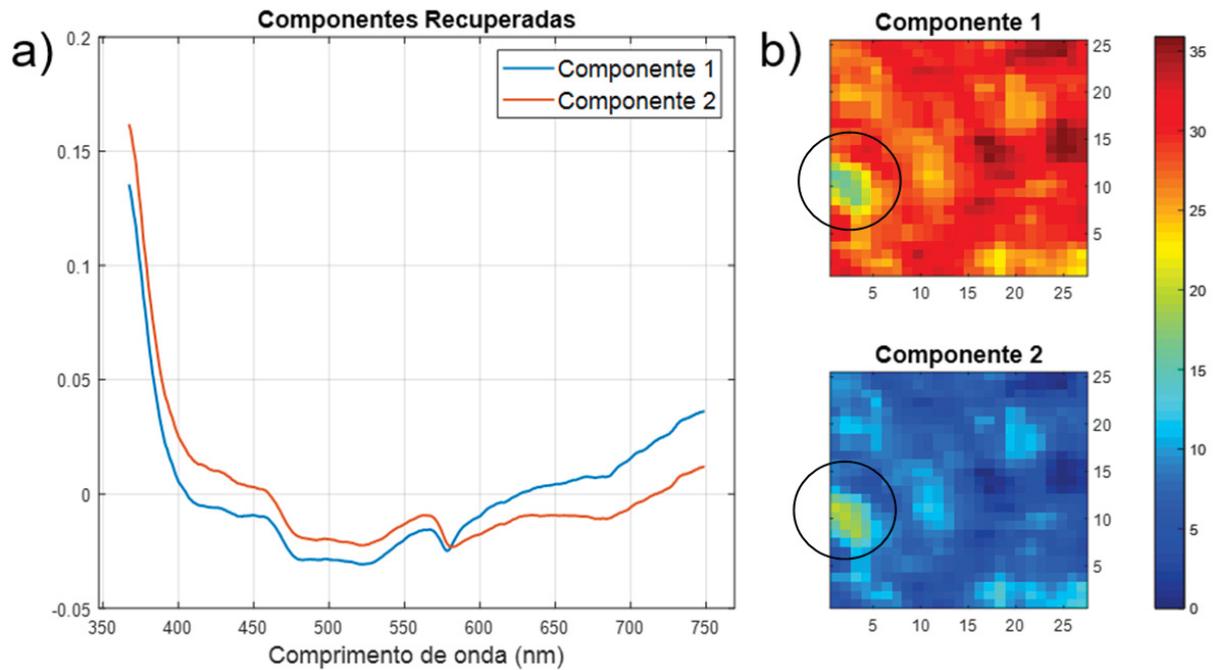
FONTE: O autor (2024).

Com isso, não há percas significativas de informação espectral ao realizar um alisamento com janela de 150 e 101.

Para exemplificar a análise de dados conduzida para cada amostra, vamos usar a mesma amostra usada para DD-SIMCA e PLS-DA composta de café Arábica e milho.

A ferramenta foi aplicada buscando encontrar as duas componentes puras dessa amostra. A Figura 31 apresenta a resposta, perfil espectral recuperado (espectros puros das espécies) e perfil da distribuição da concentração das mesmas.

Figura 31 - Perfil espectral(a) e mapa de concentração(b) recuperada pelo MCR-ALS da amostra id\_9732422, mistura de café arábica com milho, com pré-tratamento SNV.

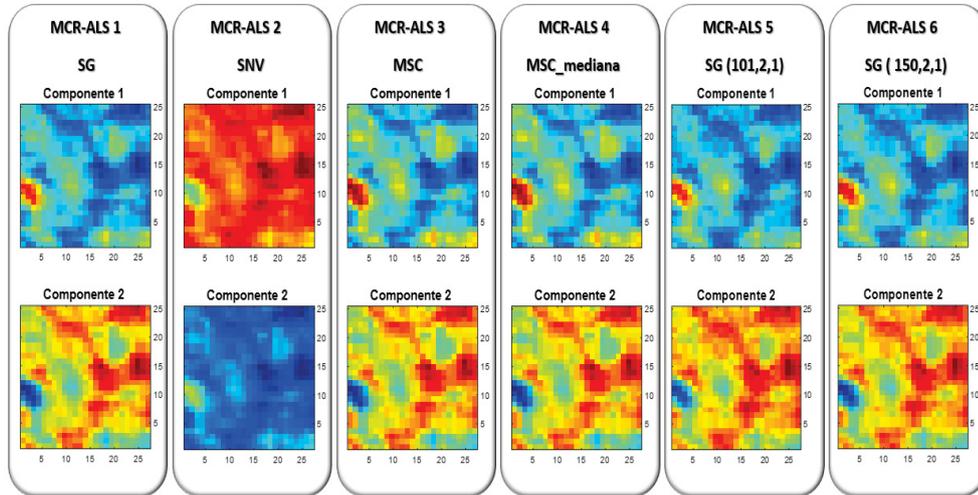


FONTE: O autor (2024).

Pela Figura 31a, temos o perfil espectral das componentes puras existentes na amostra com uma grande diferença entre eles. A Figura 31b traz a distribuição da concentração de cada uma dessas componentes no espaço do imageamento feito em 5x5mm. Cada um dos pixels dessa área tem um valor de concentração para a componente 1 e 2, que pode ser identificado pela escala de cores da figura. A componente 1 está em alta concentração ao longo de quase toda a amostra, deixando apenas um espaço em menor concentração indicado pelo círculo na imagem, justamente o mesmo espaço onde a componente 2 tem maior concentração.

Não houve diferenças significativas entre os mapas de concentração para os 6 MCR-ALS realizados. A Figura 32 apresenta o resultado para a mesma amostra de cada um dos tratamentos.

Figura 32 - Mapas de concentração recuperadas a partir de diferentes pré-tratamentos da amostra id\_9732422.



FONTE: O autor (2024).

Pode-se observar uma inversão nas componentes do MCR-ALS 2, mas isso não afeta nossa interpretação.

Com esse resultado, tem-se informações o suficiente para localizar as espécies mais distintas na amostra. Os mapas de concentração apresentam um acúmulo de concentração espacial das componentes e uma evidente separação entre elas.

O próximo passo é investigar que é cada uma dessas componentes. A atribuição de cada componente foi realizada por testes de similaridade entre os espectros recuperados e os espectros puros médios de cada um dos possíveis tipos de café e adulterantes presentes no conjunto amostral medindo o com coeficiente de correlação entre vetores ( $r^2$ ), onde quanto mais próximo de 1 o valor, maior a semelhança espectral. A Tabela 12 mostrar as atribuições para cada componente referente a amostra da Figura 31 para cada tipo de pré-tratamento.

Tabela 12 - Teste de similaridade dos espectros recuperados a partir de cada pré-tratamento.

Teste de similaridade – $r^2$						
	SG	SNV	MSC	MSC_mediana	SG(101,2,1)	SG(150,2,1)
<b>1<sup>a</sup> componente</b>	Triguilho	Arábica	Triguilho	Triguilho	Caroço de açaí	Casca
<b>2<sup>a</sup></b>	Arábica	Caroço	Arábica	Arábica	Arábica	Arábica

---

**componente** de açaí

---

Fonte: O autor (2024).

Todos os resultados recuperados de cada MCR-ALS indicaram que uma das componentes foi identificada como sendo de café arábica, sendo coerente com composição da amostra. Para o adulterante não houve acerto em nenhum dos tratamentos. Nota-se que todas componentes recuperadas tem o mesmo mapa de concentração, indicando que a variabilidade e diferença encontrar entre as componentes é clara.

Esse processo foi feito para todas as 73 amostras, 60 binárias e 13 ternárias. Os parâmetros de qualidade dos modelos (figuras de mérito) foram satisfatórios e estão apresentados na tabela abaixo.

Tabela 13 - Figura de mérito com os valores referente aos piores desempenhos para cada pré-tratamento.

	<b>SG</b>	<b>SNV</b>	<b>MSC</b>	<b>MSC_median</b>	<b>SG</b> <b>(101,2,1)</b>	<b>SG</b> <b>(150,2,1)</b>
<b>Lof (%)</b>	1,2700	2,4122	0,7197	0,7352	5,2819	3,6247
<b>R<sup>2</sup> (%)</b>	99,9839	99,9418	99,9948	99,9946	99,7190	99,8664

Fonte: O autor (2024).

Visando encontrar e identificar cada uma das componentes recuperadas para cada amostra pelos testes de similaridade, a Tabela 14 apresenta a taxa de atribuições corretas para cada tratamento em função do teste de similaridade para as amostras binárias. Para esse processo, foi aplicado um filtro, aplicando inicialmente o teste de similaridade das componentes recuperadas com os espectros de café. Uma vez encontrado o melhor valor de similaridade de uma componente com seu respectivo tipo de café, o próximo passo foi aplicar testes de similaridade entre a componente sem associação ao café com os espectros de adulterante puros. Isso foi feito para evitar a dupla associação de uma componente ao mesmo tipo de café, visto que sabemos se tratar de amostras binárias, café mais adulterante.

Tabela 14 - Eficiência em identificar corretamente as componentes recuperadas por MCR-ALS com teste de similaridade de amostras com misturas binárias.

<b>Teste de similaridade (r<sup>2</sup>)</b>						
	<b>SG</b>	<b>SNV</b>	<b>MSC</b>	<b>MSC_mediana</b>	<b>SG</b>	<b>SG</b>

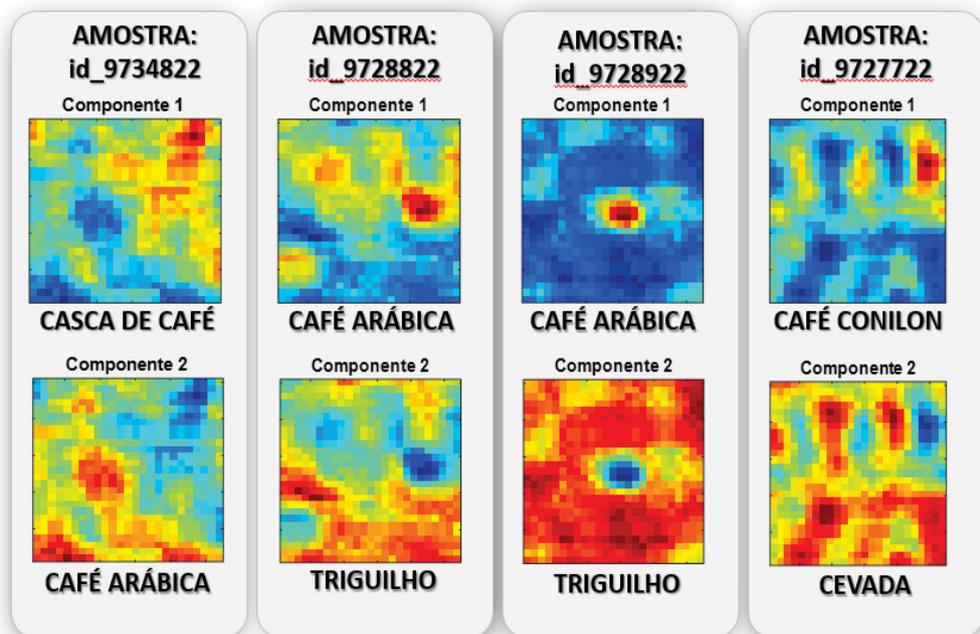
					(101,2,1)	(150,2,1)
<b>Tipos de café</b>	76,7%	80,0%	85,0%	85,0%	86,67%	93,33%
	(46/60)	(48/60)	(51/60)	(51/60)	(52/60)	(56/60)
<b>Adulterantes</b>	23,3%	18,3%	23,3%	25,0%	31,67%	30,00%
	(14/60)	(11/60)	(14/60)	(15/60)	(19/60)	(18/60)

Fonte: O autor (2024).

Pela tabela acima temos a melhor recuperação do perfil espectral relacionado a espécie de café com o pré-tratamento SG (150,2,1), atribuindo corretamente 93,33% os tipos de café que constituem a amostra. Devido à ausência de trabalhos acadêmicos com configuração similar a nossa, não temos possíveis comparações em relação a eficiência dos modelos.

Por outro lado, a identificação dos adulterantes foi bastante dificultosa e com resultados similares entre os tipos de pré-tratamentos e teste de similaridade. A Figura 33 apresenta alguns dos perfis de concentração de amostras que tiveram os adulterantes corretamente atribuídos.

Figura 33 - Mapa do perfil de concentração recuperado pelo MCR-ALS com suas respectivas componentes. Abaixo de cada mapa de cor está a atribuição de cada componente em função dos testes de similaridade.



FONTE: O autor (2024).

Nota-se que os mapas referentes aos componentes atribuídas aos adulterantes tem uma grande distribuição espacial ao longo da amostra. Isso pode

ajudar a minimizar o efeito dos subpixels, logo, há mais variabilidade em função do adulterante no conjunto de dados, facilitando a sua recuperação e resultando em maior similaridade com os espectros puros do respectivo adulterante.

Para as amostras ternárias observamos uma eficiência inferior na atribuição do tipo de café. Isso está relacionado a ausência de filtros na atribuição dos resultados em função dos testes de similaridade, como foi feito para as amostras binárias. Isso impactou na geração de resultados com atribuições duplicadas para o tipo de café.

Para avaliar desempenho das misturas terciárias, considerando apenas os MCR-ALS que geraram melhor resultado para amostras binárias. A avaliação foi feita com 3 situações em função dos resultados de similaridade: i) Identificar os dois tipos de café presente na amostra; ii) Achar alguma adulterante, seja ele qual for; iii) identificar corretamente o tipo de adulterante. A Tabela 15 apresenta a porcentagem de quantas vezes ocorrerem cada uma das situações.

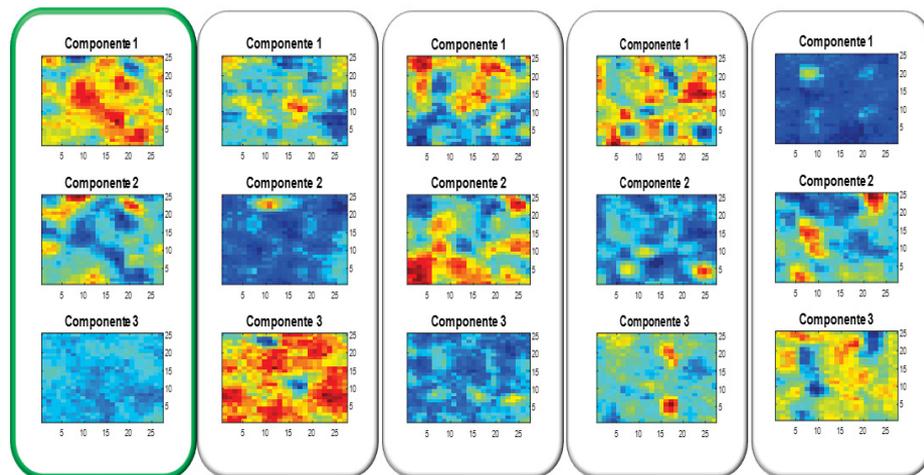
Tabela 15 - Eficiência em identificar corretamente as componentes recuperadas por MCR-ALS com teste de similaridade de amostras com misturas ternárias.

Tratamento	Identificou os dois cafés	Achou adulterante	Identificou o adulterante
SG (101,2,1)	38,46%	84,61%	38,46%
SG (150,2,1)	30,76%	84,61%	30,76%

Fonte: O autor (2024).

Entre esse grupo de amostras, apenas uma foi 100% identificada, isso é, acertar os tipos de café e o adulterante. Na Figura 34 temos mapas recuperados de 5 amostras do tratamento SG(101,2,1), e com contorno verde, temos a única amostra que foi 100% classificada. Para essa amostra, a classificação das três componentes foi Café arábica, Café conilon e Milho, respectivamente.

Figura 34 - MCR-ALS de 5 amostras de misturas terciárias.



Fonte: O autor (2024).

Apesar da baixa acurácia na identificação de todas as componentes que constituem a amostra, a ferramenta entrega com grande clareza a variabilidade existente na amostra de 5x5mm. Podemos observar com grande contraste pontos isolados na área analisada, indicando a separação espacial entre as componentes existentes na amostra. Isso já é um avanço quando comparado ao modo de catação manual feito na determinação de adulterantes em amostra de café torrado moído realizado pelo MAPA. Para mistura binária, com auxílio do teste de similaridade entre vetores, cerca de 93% o tipo de café foi identificado e localizado, com isso, o encarregado na determinação do laudo pode ser direcionado a uma área menor de análise.

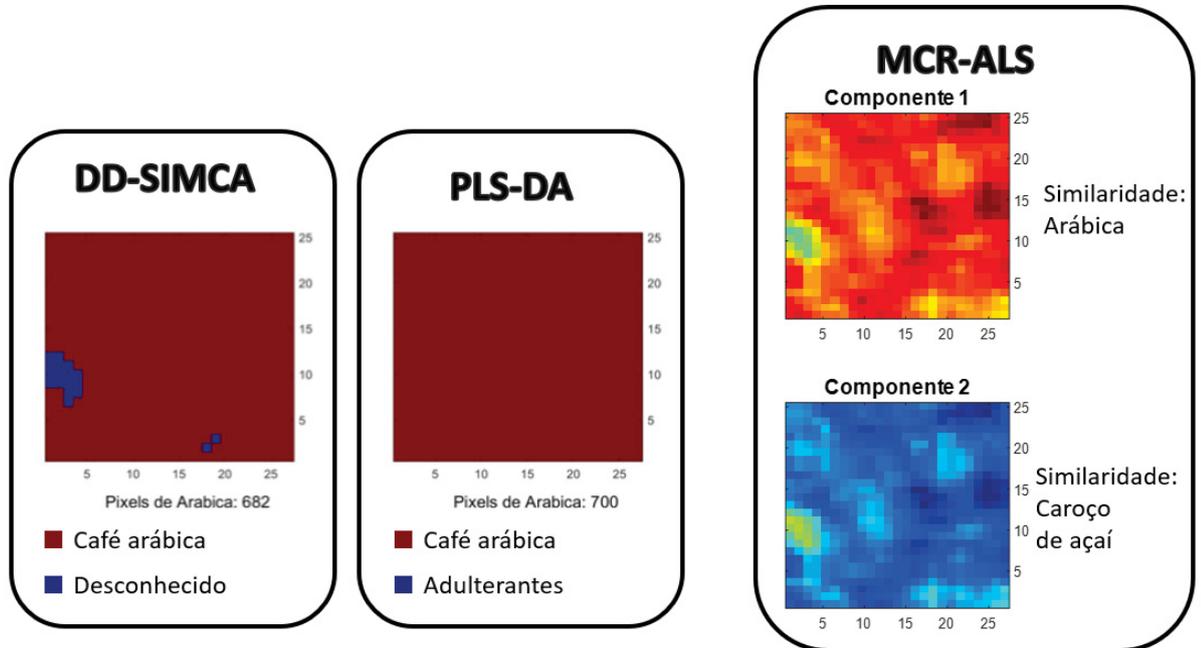
#### 5.4.4 Comparação entre os mapas recuperados: PLS-DA vs. DD-SIMCA vs. MCR-ALS

Nesse tópico, será agrupado os resultados dos modelos supervisionados e não supervisionados e comparado entre si. O pré processamento usado para comparação entre as ferramentas será SNV.

Mantendo o padrão, a Figura 35 traz o resultado da amostra de café arábica com milho apresentado em todas as ferramentas. O destaque desse resultado é como a ferramenta de MCR-ALS, um método não supervisionado consegue recuperar mapas em alta definição distinguindo espécies diferentes existentes na

amostra. Vinculado ao teste de similaridade, conseguimos identificar o tipo de café e sua localização em 96% dos casos.

Figura 35 - Mapas recuperados por teste não-supervisionados e supervisionados da amostra de café arábica mais milho.

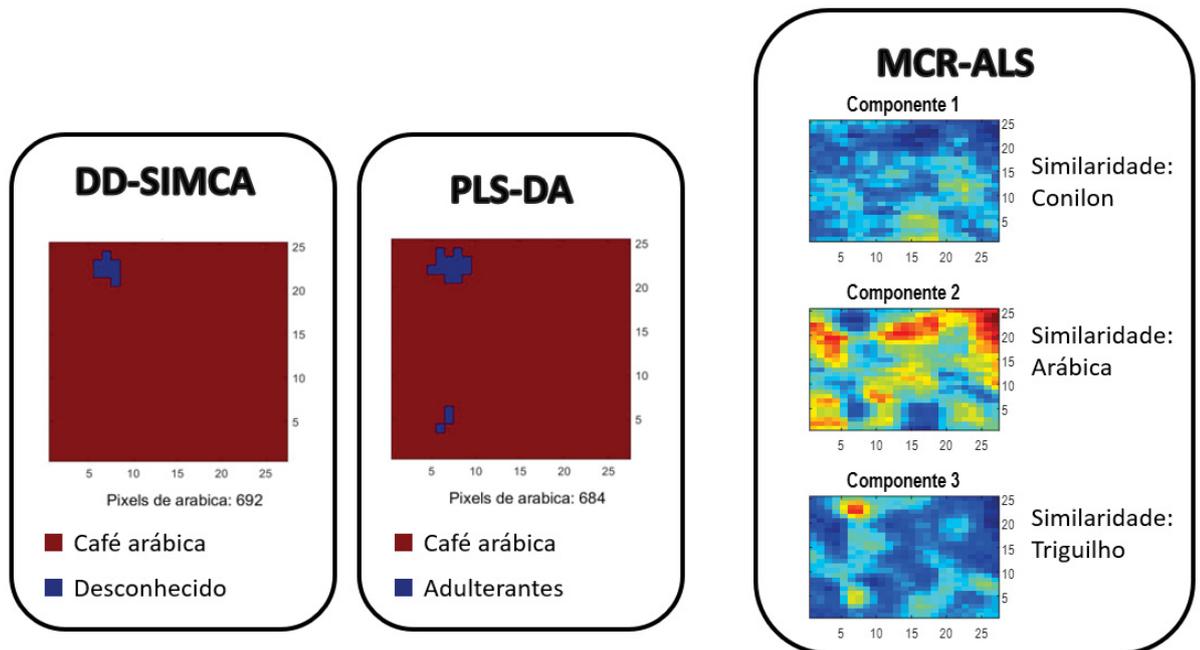


Fonte: O autor(2024).

Ao comparar com os métodos supervisionados DD-SIMCA, obtemos o mesmo mapa referente a localização do tipo de café presente na amostra e a área de não café, um possível adulterante.

Outra amostra que se destaca na identificação do adulterante é a mistura terciária de café arábica, café conilon e milho. A Figura 36 apresenta os mapas de identificação do tipo do café arábica e o MCR-ALS com o resultado do teste de similaridade.

Figura 36 - Mapas recuperados por teste não-supervisionados e supervisionados da amostra de mistura terciária.



Fonte: O autor(2024).

Os métodos supervisionados conseguem identificar a mesma área da imagem como adulterantes (ou desconhecido no caso do DD-SIMCA), com o PLS-DA atribuindo mais pontos no outro extremo da área também a adulterante. Complementando o resultado, o MCR-ALS identifica mais de um tipo de adulterante, mesmo com uma área similar aos métodos supervisionados.

No geral, o MCR-ALS com teste de similaridade para identificar suas componentes é uma ferramenta simples e eficiente para mapear as diferentes variabilidades de espectros na amostra. Vale ressaltar que para o estudo em questão tínhamos informação inicial o suficiente para assegurar o resultado final, como por exemplo a quantidade de constituintes na amostra e os tipos de adulterantes presentes.

Para uma situação de investigação de uma amostra sem nenhuma informação, as ferramentas não-supervisionada vinculada ao um simples teste de similaridade podem ser uma alternativa viável, onde não é preciso ter informação da espécie indesejável, apenas do que está procurando, visto que o MCR-ALS vinculado ao testes de correlação entre vetores associa corretamente mais de 95% das situações o tipo de café e sua localização no espaço mapeado.

## 6 CONCLUSÃO

A portaria nº 540 do MAPA, que recentemente entrou em vigor, regulamenta o controle de qualidade de café torrado moído no Brasil. E por ser uma portaria nova, os órgãos de legislação estão buscando alternativas mais simples e econômicas para propor um método de referência na detecção de adulterantes em amostras de café. Com isso, esse trabalho explorou diversos métodos e ferramentas quimiométricas aliadas a equipamentos espectrais para caracterizar e diferenciar café puro de adulterado.

O uso do equipamento NIR com auxílio da ferramenta PLS-DA e DD-SIMCA para discriminação entre os grupos café puro e café adulterado entre 3-20%(m/m) mostrou grande potencial quando aplicado ao café produzido com mesmas condições de torrefação, alcançando 100% de sensibilidade para o café puro. Porém, ao se avaliar amostras com diferentes métodos e níveis de torragem a acurácia do modelo sofreu uma diminuição.

Para o imageamento das amostras, tratando cada mapa com seu respectivo espectro médio, foi possível classificar o tipo de café majoritário dentro da amostra com sensibilidade acima de 95% para café arábica e conilon com as ferramentas PLS-DA e DD-SIMCA. Isso é especialmente interessante quando o café conilon é usado como um adulterante devido seu custo de produção inferior frente ao arábica.

Tratando os dados de imageamento pixel a pixel, observou grande coerência entre os resultados encontrados com as 3 ferramentas usadas. Com os métodos supervisionados, os modelos de café arábica para DD-SIMCA e PLS-DA encontram adulterantes nas amostras em 27,5% e 10% dos casos, respectivamente para o café arábica, e 75% e 95% para café conilon, respectivamente. Essa diferença entre os cafés pode estar associada à especificidade para cada um desses tipos de café em relação aos adulterantes, com o café arábica aproximadamente sendo 99% e para café conilon 93%.

Com o método não supervisionado, MCR-ALS e aplicando testes de similaridade espectral foi possível identificar o tipo de café e sua localização em amostras constituídas de um tipo de café mais um adulterante em 96 % dos casos na área mapeada de 5x5mm, conseqüentemente, indicando a área do possível adulterante. Além disso, estes resultados chamam a atenção, principalmente devido a facilidade em conseguir mapas de concentração das espécies puras presentes na

amostra com alta resolução espacial utilizando um sistema de imageamento de menos de R\$ 2.500,00. Caso haja a necessidade de contra-prova, este método poderia indicar ao analista a região necessária para realizar a coleta do adulterante e assim realizar uma segunda análise química, como a cromatografia líquida.

Por fim, vale ressaltar a robustez dos mapas de concentração obtidos quando se compara o 3 métodos quimiométricos empregados. E apesar de que, do ponto de vista instrumental, a faixa espectral do visível é um grande fator negativo na hora de identificar diferenças sutis entre matérias similares ao olho humano, com a alta resolução espacial, foi possível identificar pontos de diferenças e identificar elementos distintos na amostra.

## REFERÊNCIAS

1. INTERNATIONAL COFFEE ORGANIZATION (ICO). **Total production by all exporting countries.**
2. **Fairtrade and Coffee.** FAIRTRADE ed. London: [s.n.].
3. LEAL, J. G. T. Diário oficial da união. 2022 c, p. 1–10.
4. MATOS, ESMEIRE CRUZ DE MATOSINHOS, F. C. L. et al. Atlas de Microscopia. p. 48, 2012d.
5. NALLAN CHAKRAVARTULA, S. S. et al. Use of convolutional neural network (CNN) combined with FT-NIR spectroscopy to predict food adulteration: A case study on coffee. **Food Control**, v. 135, n. January, p. 108816, 2022e.
6. KARIMVAND, S. K. et al. Multivariate curve resolution-soft independent modelling of class analogy (MCR-SIMCA). **Analytica Chimica Acta**, v. 1291, n. January, p. 342205, 2024f.
7. INTERNATIONAL COFFEE ORGANIZATION (ICO). Coffee Report and Outlook (CRO). **International Coffee Organization ICO**, n. April, p. 1–39, 2023g.
8. OUR WORLD IN DATA. **Coffee bean production, 2022.**
9. NOGUEIRA, J. G. A. **Estratégias para a Cafeicultura no Brasil.** [s.l.] Editora Atlas SA, 2015i.
10. ASSOCIAÇÃO BRASILEIRA DA INDÚSTRIA DE CAFÉ. **REGULAMENTO TÉCNICO PROGRAMA DE QUALIDADE DO CAFÉ.** [s.l: s.n.].
11. LEROY, T. et al. **Genetics of coffee qualityBraz. J. Plant Physiol.** [s.l: s.n.].
12. SCHENKER, S.; ROTHGEB, T. The Roast-Creating the Beans' Signature. Em: **The Craft and Science of Coffee.** [s.l.] Elsevier Inc., 2017l. p. 245–271.
13. PIRES, F. DE C. et al. Near-infrared spectroscopy and multivariate calibration as an alternative to the Agtron to predict roasting degrees in coffee beans and ground coffees. **Food Chemistry**, v. 365, 15 Dez. 2021m.
14. SCA. **SCA Agtron Roast Color Kit.**
15. DELTA COLOR. **Delta Color.** Disponível em: <[https://www.deltacolorbrasil.com/classificacao\\_cafe\\_scaa\\_agtron.html](https://www.deltacolorbrasil.com/classificacao_cafe_scaa_agtron.html)>.
16. LEME, D. S. et al. Recognition of coffee roasting degree using a computer vision system. **Computers and Electronics in Agriculture**, v. 156, p. 312–317, 1 Jan. 2019p.
17. DE CARVALHO POLARI SOUTO, U. T. et al. Identification of adulteration in ground roasted coffees using UV-Vis spectroscopy and SPA-LDA. **Lwt**, v. 63, n. 2, p. 1037–1041, 2015q.
18. BAQUETA, M. R. et al. Brazilian Canephora coffee evaluation using NIR spectroscopy and discriminant chemometric techniques. **Journal of Food Composition and Analysis**, v. 116, n. September 2022, p. 0–2, 2023r.
19. ZHU, M. et al. Fast determination of lipid and protein content in green coffee beans from different origins using NIR spectroscopy and chemometrics. **Journal of Food Composition and Analysis**, v. 102, n. October 2020, p. 104055, 2021s.
20. AGTRON INC. **M-BASIC III.**
21. AMIGO, J. M. **Hyperspectral Imaging for Food Quality Analysis and Control.** [s.l: s.n.]. v. 54
22. LODHI, V.; CHAKRAVARTY, D.; MITRA, P. Hyperspectral Imaging System: Development Aspects and Recent Trends. **Sensing and Imaging**, v. 20, n.

1, 1 Dez. 2019v.

23. FENG, Y. Z.; SUN, D. W. Application of Hyperspectral Imaging in Food Safety Inspection and Control: A Review. **Critical Reviews in Food Science and Nutrition**, v. 52, n. 11, p. 1039–1058, Nov. 2012w.

24. WANG, J. et al. PCA-U-Net based breast cancer nest segmentation from microarray hyperspectral images. **Fundamental Research**, v. 1, n. 5, p. 631–640, 2021x.

25. PÉREZ, M. et al. Technical non-invasive study of the novo-hispanic painting the Pentecost by Baltasar de Echave Orio by spectroscopic techniques and hyperspectral imaging: In quest for the painter's hand. **Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy**, v. 250, 2021y.

26. CHAO, K.; YANG, C. C.; KIM, M. S. Spectral line-scan imaging system for high-speed non-destructive wholesomeness inspection of broilers. **Trends in Food Science and Technology**, v. 21, n. 3, p. 129–137, Mar. 2010z.

27. GOWEN, A. A. et al. **Hyperspectral imaging combined with principal component analysis for bruise damage detection on white mushrooms [Agaricus bisporus]**. Journal of Chemometrics. **Anais...**Mar. 2008aa

28. TANTINANTRAKUN, A. et al. Nondestructive evaluation of SW-NIRS and NIR-HSI for predicting the maturity index of intact pineapples. **Postharvest Biology and Technology**, v. 195, n. October 2022, p. 112141, 2023ab.

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

30. FERREIRA, M. et al. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, v. 22, p. 724–731, 1999ad.

31. GEMPERLINE, P. **Practical guide to chemometrics**. 2. ed. [s.l.] CRC/Taylor & Francis, 2006ae.

32. NOWICKA, P.; WOJDYŁO, A.; LASKOWSKI, P. Principal component analysis (PCA) of physicochemical compounds' content in different cultivars of peach fruits, including qualification and quantification of sugars and organic acids by HPLC. **European Food Research and Technology**, v. 245, n. 4, p. 929–938, 1 Abr. 2019af.

33. PEREIRA, F. M. V.; COLNAGO, L. A. Determination of the Moisture Content in Beef Without Weighing Using Benchtop Time-Domain Nuclear Magnetic Resonance Spectrometer and Chemometrics. **Food Analytical Methods**, v. 5, n. 6, p. 1349–1353, Dez. 2012ag.

34. ROCHA BAQUETA, M. et al. Integrated <sup>1</sup>H NMR fingerprint with NIR spectroscopy, sensory properties, and quality parameters in a multi-block data analysis using ComDim to evaluate coffee blends. **Food Chemistry**, v. 355, 1 Set. 2021ah.

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

36. FORINA, M.; CASALE, M.; OLIVERI, P. 4.04 Application of Chemometrics to Food Chemistry. Em: [s.l: s.n.].

37. OTTO, M. **Chemometrics: statistics and computer application in analytical chemistry**. [s.l.] John Wiley & Sons, 2016ak.

38. GHOSH, D.; CHATTOPADHYAY, P. Application of principal component analysis (PCA) as a sensory assessment tool for fermented food products. **Journal**

of **Food Science and Technology**, v. 49, n. 3, p. 328–334, Jun. 2012al.

39. MARK, H.; WORKMAN JR, J. **Chemometrics in spectroscopy**. [s.l.] Elsevier, 2010am.

40. PÉREZ-ENCISO, M.; TENENHAUS, M. Prediction of clinical outcome with microarray data: A partial least squares discriminant analysis (PLS-DA) approach.

**Human Genetics**, v. 112, n. 5–6, p. 581–592, Mai. 2003an.

41. DE SANTANA, F. B. et al. Experimento Didático De Quimiometria Para Classificação De Óleos Vegetais Comestíveis Por Espectroscopia No Infravermelho Médio Combinado Com Análise Discriminante Por Mínimos Quadrados Parciais: Um Tutorial, Parte V. **Química Nova**, v. 43, n. 3, p. 371–381, 2020ao.

42. SUHANDY, D.; YULIA, M. Peaberry coffee discrimination using UV-visible spectroscopy combined with SIMCA and PLS-DA. **International Journal of Food Properties**, v. 20, p. S331–S339, 18 Dez. 2017ap.

43. YULIA, M.; SUHANDY, D. **Indonesian palm civet coffee discrimination using UV-visible spectroscopy and several chemometrics methods**. Journal of Physics: Conference Series. **Anais...**Institute of Physics Publishing, 17 Abr. 2017aq

44. BASSBASI, M. et al. Prediction of the geographical origin of butters by partial least square discriminant analysis (PLS-DA) applied to infrared spectroscopy (FTIR) data. **Journal of Food Composition and Analysis**, v. 33, n. 2, p. 210–215, Mar. 2014ar.

45. ZONTOV, Y. V. et al. DD-SIMCA – A MATLAB GUI tool for data driven SIMCA approach. **Chemometrics and Intelligent Laboratory Systems**, v. 167, p. 23–28, 2017as.

46. POMERANTSEV, A. L.; RODIONOVA, O. Y. Concept and role of extreme objects in PCA/SIMCA. **Journal of Chemometrics**, v. 28, n. 5, p. 429–438, 2014at.

47. RODIONOVA, O. Y.; TITOVA, A. V.; POMERANTSEV, A. L. Discriminant analysis is an inappropriate method of authentication. **TrAC - Trends in Analytical Chemistry**, v. 78, p. 17–22, 2016au.

48. POMERANTSEV, A. L.; RODIONOVA, O. Y. New trends in qualitative analysis: Performance, optimization, and validation of multi-class and soft models. **TrAC - Trends in Analytical Chemistry**, v. 143, p. 116372, 2021av.

49. DE SOUZA, R. R.; FERNANDES, D. D. DE S.; DINIZ, P. H. G. D. Honey authentication in terms of its adulteration with sugar syrups using UV–Vis spectroscopy and one-class classifiers. **Food Chemistry**, v. 365, n. June, p. 1–7, 2021aw.

50. TOTARO, M. P. et al. Application of NIR spectroscopy coupled with DD-SIMCA class modelling for the authentication of pork meat. **Journal of Food Composition and Analysis**, v. 118, n. February, p. 105211, 2023ax.

51. BATISTA, L. et al. UV-Vis spectroscopy and one-class modeling for the authentication of the geographical origin of green coffee beans from Cerrado Mineiro, Brazil. **Journal of Food Composition and Analysis**, v. 123, n. July, p. 105555, 2023ay.

52. NARDECCHIA, A.; DUPONCHEL, L. Randomised SIMPLISMA: Using a dictionary of initial estimates for spectral unmixing in the framework of chemical imaging. **Talanta**, v. 217, n. April, p. 121024, 2020az.

53. MAEDER, M.; DE JUAN, A. **Two-Way Data Analysis: Evolving Factor Analysis**. Second Edi ed. [s.l.] Elsevier, 2020ba. v. 2

54. LABORDE, A. et al. Detection of chocolate powder adulteration with peanut using near-infrared hyperspectral imaging and Multivariate Curve Resolution. **Food Control**, v. 119, n. April 2020, p. 107454, 2021bb.

55. BORBA, F. D. S. L. et al. Confocal Raman imaging and chemometrics applied to solve forensic document examination involving crossed lines and obliteration cases by a depth profiling study. **Analyst**, v. 142, n. 7, p. 1106–1118, 2017bc.
56. JAUMOT, J.; TAULER, R.; JUAN, A. DE. MCR-ALS 2.0 toolbox. 7 Mar. 2022bd.
57. VAN DER MEER, F. The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. **International Journal of Applied Earth Observation and Geoinformation**, v. 8, n. 1, p. 3–17, 2006be.
58. NÚÑEZ, N.; SAURINA, J.; NÚÑEZ, O. Non-targeted HPLC-FLD fingerprinting for the detection and quantitation of adulterated coffee samples by chemometrics. **Food Control**, v. 124, 2021bf.