

UNIVERSIDADE FEDERAL DO PARANÁ

SHEILA CATARINA DE OLIVEIRA

ESTRATÉGIAS ANALÍTICAS MULTIVARIADAS APLICADAS NA
CARACTERIZAÇÃO E DETERMINAÇÃO DE ADULTERANTES EM ERVA-MATE
(Ilex paraguariensis St. Hill)

CURITIBA

2023

SHEILA CATARINA DE OLIVEIRA

ESTRATÉGIAS ANALÍTICAS MULTIVARIADAS APLICADAS NA
CARACTERIZAÇÃO E DETERMINAÇÃO DE ADULTERANTES EM ERVA-MATE
(*Ilex paraguariensis* St. Hill)

Tese apresentada ao Programa de Pós-Graduação em Química, no Setor de Exatas, na Universidade Federal do Paraná, como requisito parcial à obtenção do título de Doutora em Química.

Área de Concentração: Química Analítica
Data de admissão: 11/10/2018

Orientadora: Prof^a. Dr^a. Noemi Nagata
Coorientador: Prof. Dr. Frederico L. F. Soares

CURITIBA

2023

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

Oliveira, Sheila Catarina de
Estratégias analíticas multivariadas aplicadas na caracterização e
determinação de adulterantes em erva-mate (*Ilex paraguariensis* St. Hill). /
Sheila Catarina de Oliveira. – Curitiba, 2023.
1 recurso on-line : PDF.

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

Orientador: Noemi Nagata
Coorientador: Frederico L. F. Soares

1. Controle de qualidade. 2. Indicações geográficas (Marcas de origem). 3.
Erva-mate. I. Universidade Federal do Paraná. II. Programa de Pós-
Graduação em Química. III. Nagata, Noemi. IV Soares, Frederico L. F.. V.
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 tese de Doutorado de **SHEILA CATARINA DE OLIVEIRA** intitulada: **ESTRATÉGIAS ANALÍTICAS MULTIVARIADAS APLICADAS NA CARACTERIZAÇÃO E DETERMINAÇÃO DE ADULTERANTES EM ERVA-MATE (*Ilex paraguariensis* st. Hill)**, sob orientação da Profa. Dra. NOEMI NAGATA, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutora 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, 25 de Setembro de 2023.

Assinatura Eletrônica
26/09/2023 08:03:30.0

NOEMI NAGATA
Presidente da Banca Examinadora

Assinatura Eletrônica
26/09/2023 16:19:25.0

MARCOS ROGÉRIO MAFRA
Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
26/09/2023 10:29:22.0

ANDREA PINTO DE OLIVEIRA
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
26/09/2023 07:11:48.0

GILBERTO ABATE
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
28/09/2023 13:32:31.0

PAULO HENRIQUE GONÇALVES DIAS DINIZ
Avaliador Externo (UNIVERSIDADE FEDERAL DO OESTE DA BAHIA)

À minha família pelo apoio incondicional,
encorajamento, amor e pelos ensinamentos
que formaram os alicerces da minha história.
Ao meu esposo Luandson, pelo cuidado e
parceria, DEDICO.

AGRADECIMENTOS

Aos meus pais, Marli e Alencar, por todo o esforço que me permitiu chegar até aqui e pela compreensão por minha ausência; foi deles que vieram os ensinamentos mais importantes como respeito, dignidade e perseverança e também o amor incondicional.

À minha irmã Kemilly e ao meu sobrinho Dylan, por sempre compartilharem suas gracinhas, me fazendo rir e tornando os dias mais leves.

Ao meu esposo Luandson, obrigada por ser meu apoio nas horas mais difíceis, obrigada pelo amor, paciência, confiança e compreensão que me fortalecem e fazem querer continuar. E principalmente, obrigada por acreditar em mim mesmo quando eu não acredito.

Aos meus orientadores Prof^ª. Dr^ª. Noemi Nagata, Prof^ª. Dr^ª. Poliana Macedo dos Santos e Prof. Dr. Frederico Luis Felipe Soares, pelo incentivo, apoio, confiança, pelas conversas e por serem exemplos de profissionais. Muito obrigada!

Aos membros da banca de avaliadores desse trabalho: Professores(as) Dra. Andrea Pinto de Oliveira, Dr. Gilberto Abate, Dr. Marcos Rogério Mafra e Dr. Paulo Henrique Gonçalves Dias Diniz, pela gentileza em fazer parte dessa defesa e pelas contribuições significativas para o aprimoramento desse trabalho.

Aos professores dos Cursos de Pós-Graduação em Química da UFPR, que contribuíram, por meio de disciplinas, seminários e conversas informais com o desenvolvimento desta pesquisa e com a minha formação profissional.

À Universidade Federal do Paraná e ao Programa de Pós-graduação em Química pela disponibilização do espaço, equipamentos e materiais para a realização das análises.

À CAPES pela concessão da bolsa de estudos.

À Leice e Ana por todo o suporte na realização dos experimentos de RMN.

Ao Thyago por ser o melhor aluno de iniciação científica com quem alguém poderia trabalhar. Obrigada por tudo Thy, e conta comigo para o que precisar!

A todos os colegas e amigos do GAQQ, TECNOTRATER e DATALab, pelos momentos de amizade, descontração, discussão e encorajamento.

Às minhas amigas Sandra e a Leticia que foram apoio das mais diversas formas e que com alegria e amizade me ajudaram a trilhar esse caminho. Que bom que eu encontrei vocês!

Ao meu amigo Fábio, por nunca me deixar desistir. Foram tantas mensagens dizendo “Tu vai conseguir, migs”, que eu me convenci de que iria.

A todos que colaboraram direta ou indiretamente com este trabalho, participando e fortalecendo a construção do conhecimento, o meu muito obrigada!

“A resposta certa não importa nada:
o essencial é que as perguntas estejam certas”
(Mário Quintana)

RESUMO

A erva-mate é um produto de importância cultural e econômica na região Sul do Brasil que apresenta diversas propriedades nutricionais devido à sua composição rica em compostos bioativos. A qualidade e a percepção sensorial do produto é influenciada por diversos fatores, tais como, clima, solo manejo, processamento e embalagens, que afetam o perfil químico do produto. Nos últimos anos, as regiões produtoras de erva-mate têm buscado o reconhecimento e valorização dos seus produtos através da iniciativa de Indicações Geográficas. Entretanto, na maioria dos casos os produtos recebem o selo relacionado com a tradição de consumo e não chegam a ser identificados marcadores químicos que diferenciem os produtos de diferentes regiões. Outra limitação da cadeia produtiva da erva-mate é o controle de qualidade relacionado à presença de adulterantes no produto final. Nos últimos anos, o uso de métodos quimiométricos associados à espectroscopia e imagens digitais, têm mostrado potencial de aplicação na discriminação de amostras de acordo com a origem geográfica e também na identificação de adulterantes ou substituintes em amostras alimentícias. Com base nessas informações, esse estudo apresenta estratégias analíticas multivariadas associadas à espectroscopia MIR, NIR e de RMN ^1H para a discriminação de amostras produzidas na região de São Mateus do Sul. A PCA dos dados NIR não possibilitou a discriminação das amostras de acordo com a origem geográfica, mas permitiu a visualização de agrupamentos relacionados ao tipo de embalagem e à presença de sacarose nas amostras. O método PLS-DA foi então aplicado na classificação de amostras adoçadas, resultando em um modelo com 83,3% de sensibilidade e especificidade para as amostras adoçadas e 100% de sensibilidade e especificidade para as amostras não-adoçadas. Quando aplicada aos dados de RMN ^1H , a PCA possibilitou a visualização da discriminação das amostras de acordo com a origem geográfica. Sendo assim, um modelo PLS-DA foi construído para a identificação dos *buckets* contendo os deslocamentos químicos com maior influência na classificação. O modelo resultou em taxas de classificação corretas acima de 78% e possibilitou a identificação das regiões de δ_{H} 4,00-3,20 e δ_{H} 1,5-0,7, caracterizadas por polifenóis, açúcares, aminoácidos, ácidos graxos e saponinas como as mais significativas na discriminação das amostras. Além disso, um modelo PLS foi construído a partir dos dados obtidos via imagens digitais para a determinação da quantidade de talos nas amostras. O modelo de melhor desempenho empregou o sistema de cores HSV e apresentou resultados satisfatórios de ajuste linear e erros ($R^2_{\text{CV}}=0,88$ $R^2_{\text{P}}= 0,83$ $\text{RMSECV}= 5,75\%$ $\text{RMSEP}= 6,41\%$). Diante do fato de que atualmente não existem metodologias de análise direta estabelecidas para a discriminação de amostras de erva-mate de acordo com a origem geográfica, à nível de cidades ou microrregiões, e para a determinação de adulterantes ou substituintes, as estratégias apresentadas nesse estudo apresentam potencial de serem aplicadas para facilitar o controle de qualidade do produto.

Palavras-chave: Discriminação geográfica; Controle de qualidade; Mínimo preparo de amostras

ABSTRACT

Yerba mate is a culturally and economically significant product in the southern region of Brazil, characterized by its rich composition of bioactive compounds that are responsible for its various nutritional properties. The quality and sensory perception of the product are influenced by several factors, including climate, soil management, processing, and packaging, all of which impact the chemical profile of the product. In recent years, yerba mate-producing regions have sought recognition and appreciation for their products through Geographic Indication initiatives. However, in most cases, products receive the recognition based on consumption tradition and not based on the identification of chemical markers that can distinguish products from different regions. Another limitation in the yerba mate production chain is the quality control related to the presence of adulterants or substituents in the final product. In recent years, the use of chemometric methods combined with spectroscopy and digital imaging has shown potential for application in discriminating samples based on geographical origin and identifying adulterants or substitutes in food samples. Based on these insights, this study presents multivariate analytical strategies associated with MIR, NIR, and ^1H NMR spectroscopy for discriminating samples produced in the region of São Mateus do Sul. PCA of the NIR data did not enable discrimination of samples based on geographical origin but allowed for visualization of clusters related to packaging type and the presence of sucrose in the samples. The PLS-DA method was then applied to classify samples that had addition of sucrose, resulting in a model with 83.3% sensitivity and specificity for sweetened samples and 100% sensitivity and specificity for unsweetened samples. When applied to ^1H NMR data, PCA facilitated the discrimination of samples based on geographical origin. Consequently, a PLS-DA model was constructed to identify the *buckets* containing chemical shifts with the most significant influence on classification. The model yielded correct classification rates above 78% and enabled the identification of the regions at δ_{H} 4.00-3.20 and δ_{H} 1.5-0.7, characterized by polyphenols, sugars, amino acids, fatty acids, and saponins as the most significant in sample discrimination. Additionally, a PLS model was developed based on the digital image data to determine the quantity of stems in the samples. The best performing model employed the HSV color system and showed satisfactory results in linear fitting and errors ($R^2_{\text{CV}}=0,88$ $R^2_{\text{P}}= 0,83$ $\text{RMSECV}= 5,75\%$ $\text{RMSEP}= 6,41\%$). In light of the current absence of established direct analytical methodologies for discriminating of yerba mate samples based on their geographical origin, at municipal or microregional levels, as well as for the detection of adulterants or substitutes, the strategies presented in this study exhibit the potential to enhance the quality control process of this product.

Keywords: Geographic discrimination; Quality control; Minimal sample preparation.

LISTA DE FIGURAS

FIGURA 1 – DESTAQUE DA REGIÃO DETENTORA DO SELO IG-MATHE NO ESTADO DO PARANÁ	20
FIGURA 2 - MAPEAMENTO SIMPLIFICADO DO PROCESSO PRODUTIVO DA ERVA-MATE.....	22
FIGURA 3 - COMPARAÇÃO ENTRE A ERVA-MATE "VERDE" E A "DESCANSADA"	24
FIGURA 4 - PRINCIPAIS GRUPOS DE COMPOSTOS BIOATIVOS PRESENTES NA ERVA-MATE.....	25
FIGURA 5 - REPRESENTAÇÃO DO ACESSÓRIO DRIFT ACOPLADO NA CÉLULA (A) E DO SEU FUNCIONAMENTO GERAL (B).....	29
FIGURA 6 - EFEITO DA APLICAÇÃO DE UM CAMPO MAGNÉTICO NOS SPINS NUCLEARES.....	30
FIGURA 7 - EFEITO DA BLINDAGEM EM NÚCLEOS EXPOSTOS À DIFERENTES AMBIENTES QUÍMICOS.....	30
FIGURA 8 – REPRESENTAÇÃO DO SINAL QUÍMICO DO TETRAMETILSILANO NO RMN.....	31
FIGURA 9 - REPRESENTAÇÃO DO ÂNGULO MÁGICO (θ) ENTRE O CAMPO MAGNÉTICO (B_0) E O EIXO DE ROTAÇÃO EM UMA SONDA RMN HR-MAS	32
FIGURA 10 – MODELO DE IMAGEM DIGITAL REPRESENTADA EM ESCALA DE CINZA	34
FIGURA 11 - REPRESENTAÇÃO DOS CANAIS DE COR DO MODELO RGB.....	34
FIGURA 12 - REPRESENTAÇÃO DO ESPAÇO TRIDIMENSIONAL DO MODELO DE CORES RGB	35
FIGURA 13 - REPRESENTAÇÃO DOS CANAIS DE COR DO MODELO HSV	35
FIGURA 14 - REPRESENTAÇÃO DOS HISTOGRAMAS DE COR OBTIDOS A PARTIR DO MODELO RGB	36
FIGURA 15 - REPRESENTAÇÃO DO MÉTODO DE ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)	41
FIGURA 16 - REPRESENTAÇÃO DO GRÁFICO DE <i>LEVERAGE VS</i> RESÍDUOS DE <i>STUDENT</i> UTILIZADO PARA DETECÇÃO DE <i>OUTLIERS</i>	44
FIGURA 17 - REPRESENTAÇÃO DA MATRIZ <i>X</i> E DO VETOR <i>Y</i> UTILIZADOS NO MÉTODO PLS-DA	45

FIGURA 18 - REPRESENTAÇÃO DA MATRIZ DE CONFUSÃO E CARACTERIZAÇÃO DOS TIPOS DE ERROS DE CLASSIFICAÇÃO	46
FIGURA 19 - ETAPAS DO QUARTEAMENTO DAS AMOSTRAS DE ERVA-MATE.....	55
FIGURA 20 - APARATO INSTRUMENTAL UTILIZADO NA OBTENÇÃO DAS IMAGENS DIGITAIS	57
FIGURA 21 - ROTINA DE RECONHECIMENTO DOS POÇOS	58
FIGURA 22 - ROTINA DE APLICAÇÃO DA MÁSCARA DE IDENTIFICAÇÃO DOS POÇOS	59
FIGURA 23 - IDENTIFICAÇÃO DA POSIÇÃO DOS POÇOS E DAS REGIÕES DE INTERESSE PARA CADA AMOSTRA	59
FIGURA 24 - PROCESSO DE OBTENÇÃO DOS HISTOGRAMAS DE COR A PARTIR DA ROTINA DE RECONHECIMENTO DE POÇOS	60
FIGURA 25 - PERFIL ESPECTRAL DE AMOSTRAS DE ERVA-MATE NA REGIÃO MIR OBTIDOS COM ACESSÓRIO DRIFT.....	62
FIGURA 26 - REGIÃO DE <i>FINGERPRINT</i> MIR DAS 30 AMOSTRAS DE ERVA-MATE ANTES (A) E APÓS (B) PRÉ-PROCESSAMENTO.....	62
FIGURA 27 - AUTOVALORES EM FUNÇÃO DO NÚMERO DE PC'S PARA DADOS MIR.....	63
FIGURA 28 - GRÁFICOS DE <i>SCORES</i> (PC1 VS PC3) (A) E <i>LOADINGS</i> (B) OBTIDO A PARTIR DOS ESPECTROS MIR DAS AMOSTRAS COMERCIAIS DE ERVA-MATE IDENTIFICADAS DE ACORDO COM A ORIGEM GEOGRÁFICA E A PRESENÇA DE SACAROSE	64
FIGURA 29 - PERFIL ESPECTRAL DAS AMOSTRAS DE ERVA-MATE NA REGIÃO NIR OBTIDOS COM ACESSÓRIO DRIFT.....	67
FIGURA 30 - PERFIL ESPECTRAL NA REGIÃO NIR DAS AMOSTRAS DE ERVA-MATE APÓS PRÉ-PROCESSAMENTO (SNV+MC+SG).....	68
FIGURA 31 - VARIÂNCIA EXPLICADA EM FUNÇÃO DO NÚMERO DE PC'S PARA DADOS NIR.....	68
FIGURA 32 - GRÁFICOS DE <i>SCORES</i> (PC1 VS PC2) (A) E <i>LOADINGS</i> (B) OBTIDO A PARTIR DOS ESPECTROS NIR DAS AMOSTRAS COMERCIAIS DE ERVA-MATE IDENTIFICADAS DE ACORDO COM A PRESENÇA DE SACAROSE	70
FIGURA 33 – GRÁFICO DE <i>SCORES</i> (PC1 VS PC2) OBTIDO A PARTIR DOS ESPECTROS NIR DAS AMOSTRAS COMERCIAIS DE ERVA-MATE IDENTIFICADAS DE ACORDO COM O TIPO DE PRODUTO	71

FIGURA 34 - GRÁFICO DE <i>LOADINGS</i> (PC1 <i>VS</i> PC2) OBTIDO A PARTIR DOS ESPECTROS NIR DAS AMOSTRAS COMERCIAIS DE ERVA-MATE IDENTIFICADAS DE ACORDO COM O TIPO DE PRODUTO	71
FIGURA 35 - GRÁFICOS DE <i>SCORES</i> (PC1 <i>VS</i> PC4) OBTIDO A PARTIR DOS ESPECTROS NIR DAS AMOSTRAS COMERCIAIS DE ERVA-MATE IDENTIFICADAS DE ACORDO COM O TIPO DE EMBALAGEM.....	72
FIGURA 36 -ESPECTROS NIR DAS 30 AMOSTRAS DE ERVA-MATE COLETADOS EM DIFERENTES PERÍODOS ANTES (A) E APÓS (B) PRÉ-PROCESSAMENTO (MSC+MC+SG)	74
FIGURA 37 - GRÁFICOS DE <i>SCORES</i> (PC1 <i>VS</i> PC2) (A) E <i>LOADINGS</i> (B) OBTIDO A PARTIR DOS ESPECTROS NIR DAS AMOSTRAS COMERCIAIS DE ERVA-MATE IDENTIFICADAS DE ACORDO COM O ANO DE COLETA DOS ESPECTROS	75
FIGURA 38 - RESULTADOS DO MODELO DE CLASSIFICAÇÃO PLS-DA CONSTRUÍDO COM 5 VL'S A PARTIR DOS DADOS DA REGIÃO NIR PARA A IDENTIFICAÇÃO DE AMOSTRAS DE ERVA-MATE ADOÇADAS	77
FIGURA 39 - VALORES DA PROBABILIDADE DE CLASSIFICAÇÃO DAS AMOSTRAS DE ERVA-MATE EM CADA UMA DAS CLASSES DO MODELO PLS-DA CONSTRUÍDO PARA A IDENTIFICAÇÃO DE AMOSTRAS DE ERVA-MATE ADOÇADAS	78
FIGURA 40 - VIP SCORES DO MODELO PLS-DA CONSTRUÍDO COM 5 VL'S A PARTIR DOS DADOS DA REGIÃO NIR PARA A IDENTIFICAÇÃO DE AMOSTRAS DE ERVA-MATE ADOÇADAS	78
FIGURA 41 - ESPECTROS DE RMN HR-MAS DE ¹ H (400 MHZ) DE AMOSTRAS COMERCIAIS DE ERVA-MATE OBTIDOS COM ÁGUA (A) E METANOL (B) DEUTERADOS	80
FIGURA 42 – PERFIL METABÓLICO MÉDIO DE AMOSTRAS COMERCIAIS DE ERVA-MATE OBTIDOS VIA RMN HR-MAS DE ¹ H (400 MHZ) EM METANOL DEUTERADO.....	80
FIGURA 43 - ESQUEMA SIMPLIFICADO DAS ETAPAS DE PRÉ-PROCESSAMENTO APLICADAS AOS DADOS RMN HR-MAS DE ¹ H.....	81
FIGURA 44 – DESTAQUE DE REGIÃO ESPECTRAL ENTRE 6,0-5,0 PPM NO <i>BUCKET</i> DE 0,05 PPM	82
FIGURA 45 - VARIÂNCIA EXPLICADA <i>VERSUS</i> DO NÚMERO DE PC'S PARA DADOS DE RMN ¹ H.....	83

FIGURA 46 - GRÁFICOS DE <i>SCORES</i> (PC3 <i>VS</i> PC6) (A) E <i>LOADINGS</i> (B) OBTIDO A PARTIR DOS ESPECTROS DE RMN ¹ H DAS AMOSTRAS COMERCIAIS DE ERVA-MATE IDENTIFICADAS DE ACORDO COM A ORIGEM GEOGRÁFICA	84
FIGURA 47 - RESÍDUOS DE <i>STUDENT VS. LEVERAGE</i> : MODELO PLS-DA PARA CLASSIFICAÇÃO DE AMOSTRAS COMERCIAIS DE ERVA-MATE DE ACORDO COM A ORIGEM GEOGRÁFICA	86
FIGURA 48 - ERROS DE CLASSIFICAÇÃO <i>VS</i> VARIÁVEIS LATENTES: MODELO PLS-DA PARA CLASSIFICAÇÃO DE AMOSTRAS COMERCIAIS DE ERVA-MATE DE ACORDO COM A ORIGEM GEOGRÁFICA.....	87
FIGURA 49 - RESULTADOS DO MODELO DE CLASSIFICAÇÃO PLS-DA CONSTRUÍDO COM 6 VL'S A PARTIR DOS DADOS DE RMN ¹ H PARA A DISCRIMINAÇÃO DE AMOSTRAS DE ERVA-MATE DE ACORDO COM A ORIGEM GEOGRÁFICA.....	88
FIGURA 50 – S-PLOT DO MODELO PLS-DA CONSTRUÍDO COM 6 VL'S A PARTIR DOS DADOS DE RMN ¹ H PARA A DISCRIMINAÇÃO DE AMOSTRAS DE ERVA-MATE DE ACORDO COM A ORIGEM GEOGRÁFICA	89
FIGURA 51 - ESPECTRO DE RMN DE ¹ H (Δ_H 0,8-5,60) DA AMOSTRA B38 INDICANDO A PRESENÇA DE SINAIS CARACTERÍSTICOS PARA O ÁCIDO LINOLÊNICO ...	91
FIGURA 52 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO DIRETA ¹ H- ¹³ C (HSQC) DA AMOSTRA B38	91
FIGURA 53 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO A LONGA DISTÂNCIA ¹ H- ¹³ C (HMBC) DA AMOSTRA B38	92
FIGURA 54 - ESPECTRO DE RMN DE ¹ H (Δ_H 3,30-5,50) DA AMOSTRA B26 INDICANDO A PRESENÇA DOS CARBOIDRATOS β -GLUCOSE, α -GLUCOSE E SACAROSE..	93
FIGURA 55 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO DIRETA ¹ H- ¹³ C (HSQC) DA AMOSTRA B26	93
FIGURA 56 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO A LONGA DISTÂNCIA ¹ H- ¹³ C (HMBC) DA AMOSTRA B26	94
FIGURA 57 - ESPECTRO DE RMN DE ¹ H (Δ_H 3,20-8,00) DA AMOSTRA B06 INDICANDO A PRESENÇA DE SINAIS CARACTERÍSTICOS PARA A CAFEÍNA.....	97
FIGURA 58 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO DIRETA ¹ H- ¹³ C (HSQC) DA AMOSTRA B06	97
FIGURA 59 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO A LONGA DISTÂNCIA ¹ H- ¹³ C (HMBC) DA AMOSTRA B06	98

FIGURA 60 - ESPECTRO DE RMN DE ^1H (Δ_{H} 3,40-4,00) DA AMOSTRA B06 INDICANDO A PRESENÇA DE SINAIS CARACTERÍSTICOS PARA A TEOBROMINA	99
FIGURA 61 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO DIRETA ^1H - ^{13}C (HSQC) DA AMOSTRA B06	100
FIGURA 62 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO A LONGA DISTÂNCIA ^1H - ^{13}C (HMBC) DA AMOSTRA B06	100
FIGURA 63 - ESPECTRO DE RMN DE ^1H (Δ_{H} 6,25-7,65) DA AMOSTRA B06 INDICANDO A PRESENÇA DE SINAIS CARACTERÍSTICOS PARA O ÁCIDO CAFEICO	102
FIGURA 64 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO DIRETA ^1H - ^{13}C (HSQC) DA AMOSTRA B06	102
FIGURA 65 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO A LONGA DISTÂNCIA ^1H - ^{13}C (HMBC) DA AMOSTRA B06	103
FIGURA 66 - ESPECTRO DE RMN DE ^1H (Δ_{H} 1,85-2,40) DA AMOSTRA B38 INDICANDO A PRESENÇA DE SINAIS CARACTERÍSTICOS PARA O ÁCIDO Γ -AMINO BUTÍRICO	104
FIGURA 67 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO DIRETA ^1H - ^{13}C (HSQC) DA AMOSTRA B38	104
FIGURA 68 - ESPECTRO DE RMN DE ^1H (Δ_{H} 6,25-7,65) DA AMOSTRA B48 INDICANDO A PRESENÇA DE SINAIS CARACTERÍSTICOS PARA A QUERCETINA	105
FIGURA 69 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO DIRETA ^1H - ^{13}C (HSQC) DA AMOSTRA B48	106
FIGURA 70 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO A LONGA DISTÂNCIA ^1H - ^{13}C (HMBC) DA AMOSTRA B48	106
FIGURA 71 - HISTOGRAMAS DE COR DE AMOSTRAS DE ERVA-MATE ADICIONADAS DE TALO NA FAIXA DE 0-60% (M/M).....	108
FIGURA 72 - RESÍDUOS DE <i>STUDENT VS. LEVERAGE</i> : MODELO PLS PARA QUANTIFICAÇÃO DE TALO EM ERVA-MATE.....	110
FIGURA 73 - VALORES MENSURADOS <i>VS</i> PREDITOS: MODELO PLS PARA QUANTIFICAÇÃO DE TALO EM ERVA-MATE.....	111
FIGURA 74 - <i>VIP SCORES</i> DO MODELO PLS CONSTRUÍDO COM 3 VL'S A PARTIR DOS DADOS DE IMAGENS DIGITAIS PARA QUANTIFICAÇÃO DE TALO EM ERVA-MATE.....	111

FIGURA 75 - GRÁFICOS DE *SCORES* (PC1 *VS* PC3) OBTIDO A PARTIR DOS DADOS DE
IMAGENS DIGITAIS DAS AMOSTRAS DE ERVA-MATE IDENTIFICADAS DE
ACORDO COM A FAIXA DE CONCENTRAÇÃO DE TALO 112

LISTA DE TABELAS

TABELA 1 - IDENTIFICAÇÃO DAS AMOSTRAS DE ERVA-MATE E PRINCIPAIS INFORMAÇÕES FORNECIDAS NOS RÓTULOS.....	47
TABELA 2 – MATRIZ DE CONFUSÃO DO MODELO DE CLASSIFICAÇÃO PLS-DA CONSTRUÍDO COM 5 VL’S A PARTIR DOS DADOS DA REGIÃO NIR PARA A IDENTIFICAÇÃO DE AMOSTRAS DE ERVA-MATE ADOÇADAS.....	73
TABELA 3 – MATRIZ DE CONFUSÃO DO MODELO DE CLASSIFICAÇÃO PLS-DA CONSTRUÍDO COM 11 VL’S A PARTIR DOS DADOS DE RMN ¹ H PARA A DISCRIMINAÇÃO DE AMOSTRAS DE ERVA-MATE DE ACORDO COM A ORIGEM GEOGRÁFICA.....	84
TABELA 4 - DADOS DE RMN HR-MAS DE ¹ H E RMN 2D PARA O ÁCIDO LINOLÊNICO.....	87
TABELA 5 - DADOS DE RMN HR-MAS DE ¹ H E RMN 2D PARA A β-GLUCOSE, α-GLUCOSE E SACAROSE.....	91
TABELA 6 - DADOS DE RMN HR-MAS DE ¹ H E RMN 2D PARA A CAFEÍNA.....	92
TABELA 7 - DADOS DE RMN HR-MAS DE ¹ H E RMN 2D PARA A TEOBROMINA.....	95
TABELA 8 - DADOS DE RMN HR-MAS DE ¹ H E RMN 2D PARA O ÁCIDO CAFEICO...	97
TABELA 9 - DADOS DE RMN HR-MAS DE ¹ H E RMN 2D PARA O ÁCIDO Γ-AMINOBTÚRICO.....	99
TABELA 10 - DADOS DE RMN HR-MAS DE ¹ H E RMN 2D PARA A QUERCETINA....	103
TABELA 11 - PARÂMETROS DE DESEMPENHO DOS MODELOS DE REGRESSÃO PLS PARA A QUANTIFICAÇÃO DE TALOS EM AMOSTRAS DE ERVA-MATE COMERCIAL A PARTIR DE IMAGENS DIGITAIS.....	87
TABELA 12 – MÉDIA E DESVIO-PADRÃO DOS VALORES PREDITOS DE % DE TALO NAS AMOSTRAS DO CONJUNTO DE VALIDAÇÃO.....	89

LISTA DE ABREVIATURAS

ANVISA	Agência Nacional de Vigilância Sanitária
BPA	Boas Práticas Agrícolas
CEPEA	Centro de Estudos Avançados em Economia Aplicada
CNA	Confederação da Agricultura e Pecuária do Brasil
Coshift	<i>Correlation-shifting</i>
COSY	COrrelation SpectroscopY
COW	<i>Correlation Optimized Warping</i>
DERAL	Departamento de Economia Rural
DO	Denominação de Origem
DRIFT	<i>Diffuse Reflectance Infrared Fourier Transform</i>
DSG	Derivada <i>Savitzky–Golay</i>
FIR	<i>Far Infrared</i>
FN	<i>Falsos Negativos</i>
FP	<i>Falsos Positivos</i>
HCA	<i>Hierarchical Clustering Analysis</i>
HMBC	<i>Heteronuclear Multiple Bond Correlation</i>
HR-MAS	<i>High Resolution Magic Angle Spinning</i>
HSI	<i>Hue Saturation Intensity</i>
HSL	Hue Saturation, Lightness
HSQC	<i>Heteronuclear Single Quantum Correlation</i>
HSV	<i>Hue Saturation Value</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
Icoshift	<i>Interval correlation-shifting</i>
IG	Indicação Geográfica
INPI	Instituto Nacional de Propriedade Industrial
IP	Indicação de Procedência
IV	Infravermelho
LDA	<i>Linear Discriminan Analysis</i>
LV	<i>Latent Variable</i>
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
MC	<i>Mean Center</i>

MCR	<i>Multivariate Curve Resolution</i>
MIR	<i>Mid Infrared</i>
MLR	<i>Multiple Linear Regression</i>
MSC	Multiplicative Scatter Correction
NIR	<i>Near Infrared</i>
PARAFAC	<i>PARAllel FACtor analysis</i>
PC	<i>Principal Component</i>
PCA	<i>Principal Component Analysis</i>
PIB	Produto Interno Bruto
<i>Pixel</i>	<i>Picture Element</i>
PLS	<i>Partial Least Squares</i>
PLS-DA	<i>Partial Least Square-Discriminant Analysis</i>
PN	Padrão Nacional
PNG	<i>Portal Network Graphic</i>
RGB	<i>Red Green Blue</i>
RMN	Ressonância Magnética Nuclear
RMSEC	<i>Root Mean Square Error of Calibration</i>
RMSECV	<i>Root Mean Square Error of Cross-validation</i>
RMSEP	<i>Root Mean Square Error of Prediction</i>
ROI	<i>Region of interest</i>
SEBRAE	Serviço Brasileiro de Apoio às Micro e Pequenas Empresas
SG	<i>Suavização Savitzky–Golay</i>
SIMCA	<i>Soft Independent Modelling of Class Analogies</i>
SNV	<i>Standard Normal Variate</i>
SPA	Successive Projections Algorithm
TMS	Tetrametilsilano
VIP	<i>Variable Importance in Projection</i>
VL	Variáveis Latentes
VN	<i>Verdadeiros Negativos</i>
VP	<i>Verdadeiros Positivos</i>

SUMÁRIO

1 INTRODUÇÃO	18
2 FUNDAMENTAÇÃO TEÓRICA	19
2.1 ERVA-MATE: CONTEXTO HISTÓRICO E IMPORTÂNCIA ECONÔMICA	19
2.2 SISTEMA DE INDICAÇÕES GEOGRÁFICAS	20
2.3 PROCESSO PRODUTIVO DA ERVA-MATE	21
2.4 COMPOSIÇÃO QUÍMICA DA ERVA-MATE	24
2.5 FRAUDES EM ALIMENTOS	26
2.6 MÉTODOS ESPECTROSCÓPICOS	27
2.6.1 <i>Espectroscopia na região do infravermelho</i>	27
2.6.2 <i>Espectroscopia de ressonância magnética nuclear</i>	29
2.7 MÉTODOS BASEADOS EM IMAGENS DIGITAIS	33
2.8 QUIMIOMETRIA	36
2.8.1 <i>Organização dos dados</i>	37
2.8.2 <i>Pré-processamento dos dados</i>	38
2.8.2.1 <i>Centragem dos dados na média</i>	38
2.8.2.2 <i>Escalamento</i>	38
2.8.2.3 <i>Correção multiplicativa de espalhamento e padronização normal de sinal</i>	39
2.8.2.4 <i>Técnicas de alisamento (suavização)</i>	39
2.8.2.5 <i>Alinhamento de sinais espectrais</i>	39
2.8.3 <i>Análise de componentes principais (PCA)</i>	40
2.8.5 <i>Regressão pelo método de mínimos quadrados parciais (PLS)</i>	41
2.8.6 <i>Análise discriminante pelo método de mínimos quadrados parciais (PLS-DA)</i>	44
2.9 APLICAÇÕES DE QUIMIOMETRIA NA ANÁLISE DE ALIMENTOS	46
3 OBJETIVOS	49
3.1 OBJETIVO GERAL	49
3.2 OBJETIVOS ESPECÍFICOS.....	49
4 MATERIAL E MÉTODOS	50
4.1 AMOSTRAS.....	50
4.2 PROCESSO DE AMOSTRAGEM.....	55
4.3 AQUISIÇÃO DOS ESPECTROS NA REGIÃO DO INFRAVERMELHO.....	55
4.4 AQUISIÇÃO DOS ESPECTROS DE RMN HR-MAS DE 1H	56
4.5 AQUISIÇÃO DAS IMAGENS DIGITAIS	57

4.5.1	<i>Processamento das imagens digitais</i>	57
4.5.1.1	<i>Correção da luminosidade</i>	58
4.5.1.2	<i>Segmentação e individualização</i>	58
4.4.1.3	<i>Obtenção dos histogramas de cor</i>	59
3.7.3	<i>Preparo das misturas com adição de talos</i>	60
5	RESULTADOS E DISCUSSÃO	61
5.1	QUIMIOMETRIA APLICADA À DADOS DE ESPECTROSCOPIA MIR E NIR.....	61
5.1.1	<i>Análise exploratória comparativa – regiões MIR e NIR</i>	61
5.1.1.1	<i>Infravermelho médio (MIR)</i>	61
5.1.1.2	<i>Infravermelho próximo (NIR)</i>	66
5.1.2	<i>Análise exploratória via PCA – região NIR</i>	66
5.1.3	<i>Avaliação da influência do tempo de armazenamento de amostras de erva-mate via PCA</i>	73
5.1.4	<i>Identificação de amostras adoçadas de erva-mate via PLS-DA – região NIR</i>	75
5.2	QUIMIOMETRIA APLICADA À DADOS DE ESPECTROSCOPIA DE RMN.....	79
5.2.1	<i>Seleção do solvente deuterado</i>	79
5.2.2	<i>Análise do perfil metabólico obtido via RMN HR-MAS de ¹H</i>	79
5.2.3	<i>Análise exploratória via PCA – RMN HR-MAS de ¹H</i>	81
5.2.4	<i>Classificação das amostras de erva-mate de acordo com a origem geográfica via PLS-DA – região de RMN</i>	85
5.2.5	<i>Elucidação estrutural de marcadores via RMN bidimensional</i>	89
5.2.5.1	<i>Identificação estrutural do ácido linolênico (1)</i>	89
5.2.5.2	<i>Identificação estrutural da β-glucose (2), α-glucose (3) e sacarose (4)</i>	92
5.2.5.3	<i>Identificação estrutural da cafeína (5)</i>	95
5.2.5.4	<i>Identificação estrutural da teobromina (6)</i>	98
5.2.5.5	<i>Identificação estrutural do ácido cafeico (7)</i>	101
5.2.5.6	<i>Identificação estrutural do aminoácido GABA (8)</i>	103
5.2.5.7	<i>Identificação estrutural da quercetina (9)</i>	105
5.3	QUIMIOMETRIA APLICADA À ANÁLISE DE IMAGENS DIGITAIS.....	107
5.3.1	<i>Quantificação de talos em amostras de erva-mate via regressão PLS</i>	107
6	CONCLUSÕES	113
	REFERÊNCIAS	115

1 INTRODUÇÃO

A erva-mate (*Ilex paraguariensis* St. Hil.) é um produto florestal não-madeireiro pertencente à família das aquifoliáceas que ocorre naturalmente na América do Sul (RZASADURAN et al., 2022; GERBER et al., 2022; SCHINDLER et al., 2018). Tanto a produção como o consumo da erva-mate estão atrelados a cultura dessas regiões, especialmente aos estados do Sul do Brasil, onde é um dos principais produtos agrícolas (CEPEA, 2023).

A erva-mate é consumida nas formas de chimarrão (infusão quente) e tererê (infusão fria). Entretanto, devido à composição química complexa e às propriedades nutricionais, a planta vem sendo utilizada no desenvolvimento de novos produtos, como alimentos e cosméticos (RIBEIRO et al., 2023; VASCONCELLOS, FRAZZON & NOREÑA, 2022).

Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE, 2021), o Paraná é o maior produtor de erva-mate no país, destacando-se a cidade de São Mateus do Sul, que conta com 50% de ervais nativos ainda preservados. Dados do instituto mostram que em 2021 o Paraná foi responsável pela produção de 238 mil toneladas de erva-mate, sendo que desse total 48 mil toneladas foram produzidas na região de São Mateus do Sul.

Em 2017, a erva-mate produzida em São Mateus do Sul recebeu o selo de Indicação de Procedência como reconhecimento da história, tradição e notoriedade do produto. Apesar de ser motivo de orgulho para a região, o selo tem o potencial de estimular produtores de outras regiões a utilizar o nome da cidade a fim de obter benefícios financeiros. Outra preocupação é a adulteração do produto, seja pela adição de talos para redução de custo ou a adição de sacarose para simular o sabor mais suave e adocicado reportado com o produto de São Mateus do Sul.

A importância econômica e social do setor produtivo da erva-mate no Brasil resultou em vários estudos com o objetivo de identificar fatores que possam influenciar significativamente as propriedades químicas e sensoriais da mesma. Entretanto, estudos focados na discriminação de uma região produtora específica e o uso de métodos baseados em imagens digitais para a identificação de adulterações ainda são escassos na literatura científica.

Nesse contexto, o presente estudo se propõe a aplicação de técnicas espectroscópicas associadas às metodologias quimiométricas para a discriminação de amostras de erva-mate de acordo com a origem geográfica e a presença de sacarose. Além disso, objetiva-se o desenvolvimento de metodologias baseada em imagens digitais obtidas com um scanner de mesa para a quantificação de talos em amostras comerciais de erva-mate.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 ERVA-MATE: CONTEXTO HISTÓRICO E IMPORTÂNCIA ECONÔMICA

A história da erva-mate (*Ilex paraguariensis* St. Hill) está intrinsecamente ligada à dos povos indígenas que habitavam a América do Sul, mais especificamente na área de ocorrência natural da planta que engloba territórios do Paraguai, Argentina e a região Sul do Brasil (PIRES et al., 2016; DE OLIVEIRA & ROTTA, 1995).

Desde antes da chegada dos europeus na América, os indígenas consumiam o produto na forma de infusão ou mascada devido às propriedades estimulantes e digestivas, mas o uso foi proibido pelos jesuítas que acreditavam que a erva possuía propriedades alucinógenas (RZASA-DURAN et al., 2022). Ao invés de inibir o uso, a proibição aumentou o interesse e ajudou a difundir o consumo da erva, tanto que a comercialização do produto acabou servindo como base econômica para as missões jesuítas (MARCELO, 2014).

A produção e a comercialização da erva-mate continuaram a crescer, desempenhando um papel econômico cada vez mais importante no Brasil. O país é líder na produção de erva-mate, com uma produção média em torno de 900 mil toneladas/ano (distribuídas entre ervais nativos e plantados) responsáveis por suprir o mercado interno e as exportações, que tem como principal destino o Uruguai (DERAL, 2021).

O Estado do Paraná é responsável por aproximadamente 87% da produção da erva-mate proveniente de ervais nativos ou sombreados (DERAL, 2021). Nesse cenário, se destaca a região de São Mateus do Sul, conhecida como a capital da erva-mate, e que possui mais de 4 mil estabelecimentos agrícolas produtores da erva (IBGE, 2021).

O consumo da erva-mate, que sempre foi parte essencial da economia, cultura e identidade da região, se tornou ainda mais relevante quando em 2017 a região de São Matheus, formada por 6 municípios (FIGURA 1), recebeu o selo de indicação geográfica concedido pelo Instituto Nacional da Propriedade Industrial (INPI, 2021).

Atualmente, 29 produtores da região são credenciados para o fornecimento de erva-mate para as 7 marcas que beneficiam o produto. Para assegurar o credenciamento na Associação que recebeu o nome de IG-Mathe, os produtores precisam seguir o manual de Boas Práticas Agrícolas (BPA) que garante não só a rastreabilidade do produto, mas também que a erva será cultivada sem a utilização de defensivos agrícolas e que os produtores utilizarão métodos sustentáveis para preservar a mata nativa. O cumprimento dos processos listados no manual de especificações técnicas é fiscalizado pela própria Associação (IG-Mathe, 2023).

FIGURA 1 – DESTAQUE DA REGIÃO DETENTORA DO SELO IG-MATHE NO ESTADO DO PARANÁ



FONTE: A autora (2023).

2.2 SISTEMA DE INDICAÇÕES GEOGRÁFICAS

No passado, o consumo de alimentos baseava-se nos recursos locais, permitindo o estabelecimento quase óbvio de uma correlação entre as características de um produto e as características da região geográfica. No entanto, a globalização dos mercados de alimentos afetou esse processo, permitindo que as pessoas comam e bebam uma variedade de produtos não locais, fato que dificulta a rastreabilidade e a agregação de valor de produtos alimentícios (DIGGELMANN, 2021; CARRERA & GALLARDO, 2017; LUYKX & VAN RUTH, 2008).

Atualmente, um dos maiores desafios enfrentados pelas empresas é o de encontrar uma forma de diferenciar o seu produto dos já existentes no mercado. Um produto que apresente propriedades distintas dos seus similares tende a ser reconhecido pelos consumidores, o que eleva o nível de competitividade no comércio nacional e internacional (PINTO et al., 2020; SEBRAE, 2014). A possibilidade de identificação de um produto ou serviço como oriundo de um local, região ou país, através da análise de suas características ou parâmetros de qualidade é denominada de Indicação Geográfica (IG) (INPI, 2023).

O sistema de IG para produtos alimentícios é reconhecido mundialmente e tem como objetivos: proteger a região produtora e agregar valor ao produto (SEBRAE, 2014). No Brasil, o Instituto Nacional de Propriedade Industrial (INPI) concede duas modalidades de IG: a Indicação de Procedência (IP) e a Denominação de Origem (DO) (INPI, 2023):

Indicação de procedência: refere-se ao nome geográfico do local que se tornou conhecido como centro de extração, produção ou fabricação do produto. Nesse caso, são levados em conta a discriminação do território, os métodos de produção e a história do produto na região, e outros aspectos relacionados à tradição do consumo e comércio.

Denominação de origem: refere-se ao nome do local geográfico, que passou a designar o produto, cujas qualidades e/ou características podem ser associadas ao meio geográfico, incluindo fatores naturais e humanos. Nesse caso, análises que comprovem a correlação entre a característica e o meio geográfico devem ser apresentados.

A introdução do termo no Brasil ocorreu em 1996 na publicação da Lei de Propriedade Industrial nº 9279 e até o ano de 2023 foram concedidos pelo INPI um total de 24 selos de DO a produtos nacionais e 81 selos de IP, dentre os quais se destaca a IP conferida à erva-mate produzida na região de São Mateus do Sul. O selo impede que o nome da região seja utilizado indevidamente e com isso aumenta a garantia de que os padrões de qualidade serão mantidos, o que irá repercutir na comercialização do produto (INPI, 2023; SEBRAE, 2016).

Diversos trabalhos indicam que as propriedades sensoriais (sabor suave e adocicado) da erva-mate proveniente da região de São Matheus, são resultado da formação de diferentes compostos químicos relacionados com as características edafoclimáticas e de cultivo da região, principalmente ao fato de os ervais serem cultivados ou nativos de regiões sombreadas (DA SILVA & ROSA, 2022; MEYER et al., 2006; RAKOCEVIC et al., 2007). Apesar de indicativas, tais informações não são suficientes para comprovar que o produto de São Matheus apresente perfil químico diferenciado dos demais existentes no mercado nacional.

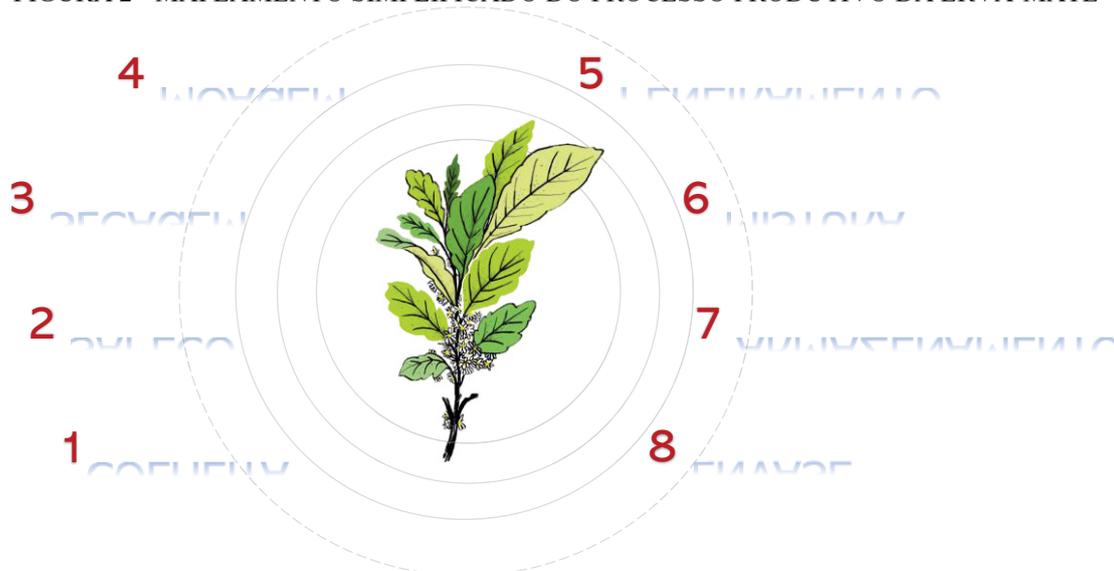
A composição química de produtos de origem vegetal, como a erva-mate, pode sofrer influência de vários fatores como: tipo de solo, clima, nível de sombreamento, diferenças no cultivo, processamento e no armazenamento (ALMEIDA et al., 2023; FERRERA et al., 2016; ROSSA et al., 2017). Sendo assim, a padronização no processo produtivo da erva produzida na região da IG, pode resultar em maior homogeneidade do perfil químico facilitando o estabelecimento da correlação entre as características do produto e o local de produção.

2.3 PROCESSO PRODUTIVO DA ERVA-MATE

Por ser uma espécie nativa do Brasil, aproximadamente metade do volume de erva-mate produzido provém do extrativismo (coleta das folhas em áreas de ocorrência natural), sendo o cultivo intencional da cultura responsável pelo restante da produção (IBGE, 2021). Os ervais nativos crescem sombreados pela Mata Atlântica e a ação humana se limita a remoção da vegetação rasteira e ao cultivo pelo método de adensamento, que consiste no plantio de mudas intercaladas com os ervais nativos. Por outro lado, os ervais cultivados são caracterizados pelo manejo convencional, que normalmente se dá à pleno sol e no qual são variadas as condições de solo, clima, topografia e vegetação natural (POLIDORO, 2022; GOULART et al., 2020).

O processamento da erva-mate (FIGURA 2) se inicia com a colheita que ocorre em dois períodos, a safra e a safrinha. A safra ocorre nos meses de maio a setembro, período em que as folhas estão maduras e que a planta está em repouso fisiológico (MACCARI JR et al., 2006). Já a safrinha ocorre no verão, entre dezembro e fevereiro, e apresenta um rendimento inferior, devido ao maior teor de água e menor espessura das folhas (DANIEL, 2009).

FIGURA 2 - MAPEAMENTO SIMPLIFICADO DO PROCESSO PRODUTIVO DA ERVA-MATE



FONTE: A autora (2023).

As principais etapas de processamento da erva-mate podem ser divididas em dois ciclos: o cancheamento e o beneficiamento. O ciclo do cancheamento inclui a colheita de folhas e talos, o sapeco e a secagem. No ciclo de beneficiamento, a erva passa por novos processamentos, como a trituração, peneiramento e mistura para a elaboração dos produtos de erva-mate (MACCARI JR et al., 2006; DE OLIVEIRA & WAQUIL, 2015).

Após o corte, as folhas da erva-mate continuam seu processo respiratório, consumindo glicose e sofrendo a ação das enzimas peroxidase e polifenoloxidase que são responsáveis por catalisar reações de oxidação dos compostos fenólicos, causando o escurecimento enzimático característico do processo de degradação do material vegetal (NABECHIMA et al., 2014). Para controlar o escurecimento e reduzir o teor de umidade, as folhas passam pelo processo de sapeco (branqueamento térmico), onde são colocadas em contato direto com a chama obtida pela combustão de toras de madeira (WENDLING, ARANTES & HELM, 2022).

Na sequência, o material vegetal segue para a etapa de secagem, que tem por objetivo aumentar a estabilidade do produto através da redução do teor de umidade para aproximadamente 5% (VALDUGA, FINZER & MOSELE, 2003, MACCARI JR, 2005). Essa

etapa é bastante relevante para a qualidade sensorial do produto final, pois várias reações químicas, tais como as reações de Maillard, ocorrem causando alterações na coloração e levam à formação de uma ampla gama de compostos, como aldeídos, cetonas e ácidos orgânicos, que contribuem para o desenvolvimento dos aromas e sabores complexos que são característicos da erva-mate (BENDLIN, 2003; OLIVEIRA et al., 2011; SAATH et al., 2012).

O material vegetal desidratado passa então para a etapa de moagem, onde a erva é socada por pilões, conhecidos como soques, pelo tempo que for necessário para atingir a granulometria desejada para o produto final (fina, média, grossa ou extra grossa). Essa etapa encerra o ciclo de cancheamento da erva, que recebe esse nome devido ao tipo de equipamento que era utilizado para moer a erva após a secagem (DANIEL, 2009).

A erva cancheada segue para o ciclo de beneficiamento, que se inicia com o peneiramento, para que se obtenha um produto homogêneo e com a quantidade adequada de folha e talos. Apesar da legislação sobre a classificação de amostras comerciais de erva-mate de acordo com a quantidade de talos presente não estar mais vigente, ainda é prática comum a indicação do padrão nacional (PN) nas embalagens. Segundo essa classificação, a erva pode ser agrupada em três categorias: PN-1, quando apresenta no mínimo 70% de folhas e no máximo 30% de outras partes do ramo; PN-2, quando apresenta no mínimo 60% de folhas e o restante de outras partes da planta; e PN-3 quando possui no mínimo 50% de folhas (BORGES, LAZZARI & LAZZARI, 2003, BRASIL, 1998).

No beneficiamento, a erva-mate ainda pode ser categorizada de acordo com o tempo de armazenamento empregado antes do envase. O produto pode passar pelo processo de armazenamento, ou maturação, de até dois anos, no qual as reações de oxidação promovem a alteração das características sensoriais da erva (MACCARI JR et al., 2005). Enquanto o mercado interno prefere o produto verde e menos amargo, com armazenamento máximo de 1 ano, o mercado externo busca produtos com maior tempo de maturação (erva-mate descansada) e que apresentam coloração amarelo-dourada e sabor mais amargo (FIGURA 3).

A erva beneficiada pode ser utilizada na formulação de produtos compostos, recebendo a adição de açúcar ou de outras espécies vegetais (por exemplo, menta e camomila). A comercialização desses compostos é regulamentada pela Agência Nacional de Vigilância Sanitária (ANVISA), mas o produto não recebe o selo de IG, visto que suas características sensoriais são alteradas pela adição de outras espécies vegetais ou sacarose (BRASIL, 2022).

FIGURA 3 - COMPARAÇÃO ENTRE A ERVA-MATE "VERDE" E A "DESCANSADA"



FONTE: PINTARELLI, 2020.

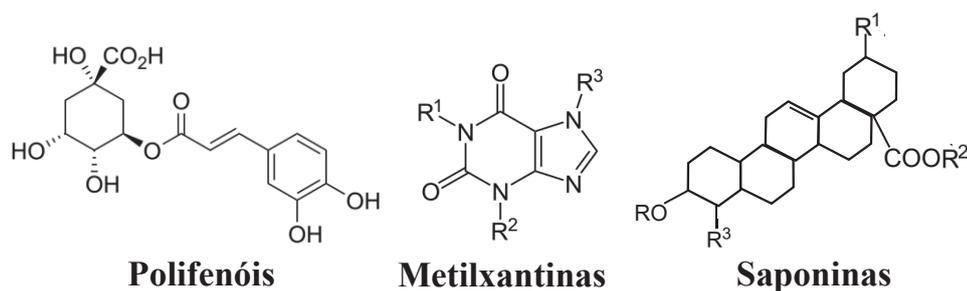
Apesar dos avanços em termos da padronização dos processos de cultivo e beneficiamento da erva-mate, a influência dos fatores edafoclimáticos na entrega sensorial e no perfil químico dos produtos ainda é significativa. Estudos reportam as diferenças encontradas em termos de cafeína, açúcares e polifenóis em amostras de erva-mate produzidas em diferentes locais ou sob diferentes condições de cultivo (DE OLIVEIRA & MARCA, 2020; DORTZBACH et al., 2020). Entretanto, devido à complexidade química da composição da erva-mate e o número de fatores que podem influenciar essa composição, a avaliação do perfil químico tem maior potencial, quando comparada a análise univariada de compostos, de propiciar a discriminação das amostras de acordo com a origem geográfica.

2.4 COMPOSIÇÃO QUÍMICA DA ERVA-MATE

Apesar da maior parte da erva-mate produzida ser destinada ao consumo em forma de infusões, o uso do produto na indústria alimentícia, farmacêutica e cosmética vem crescendo devido às suas propriedades nutricionais e medicinais, tais como, ação antioxidante, anti-inflamatório, diurética e aumento da tolerância a glicose (RIBEIRO et al., 2023; MAIZTEGUI et al., 2023; GERBER et al., 2023; FENOGLIO et al., 2021).

Essas propriedades estão intrinsicamente relacionadas com os compostos químicos encontrados na erva-mate, dentre os quais, os polifenóis, as metilxantinas e as saponinas (FIGURA 4) que são considerados os compostos bioativos mais importantes presentes nesse material vegetal (POLIDORO et al., 2023).

FIGURA 4 - PRINCIPAIS GRUPOS DE COMPOSTOS BIOATIVOS PRESENTES NA ERVA-MATE



FONTE: A autora (2023).

Os polifenóis são um grupo de compostos químicos encontrados em uma variedade de plantas, incluindo frutas, grãos, chás, e também na erva-mate. Os polifenóis podem ser divididos em várias subclasses, incluindo flavonoides, ácidos fenólicos, lignanas e estilbenos, entre outras (PRABHU et al., 2021). Cada subclasse possui diferentes estruturas químicas e propriedades, porém, esses compostos são conhecidos pelas propriedades antioxidantes e têm sido associados a uma série de benefícios à saúde (VASCONSELLOS et al., 2022).

Em estudos comparativos, a erva-mate apresentou teores elevados de compostos fenólicos quando comparados à extratos aquosos de outros tipos de chá. Segundo Marques e Farah (2009), as infusões de erva-mate apresentaram 40 vezes mais ácido 5-cafeoilquínico do que os produzidos com as infusões de *Camellia sinensis*, além de apresentar os maiores teores de ácido caféico. De acordo com Khaksar et al. (2023), as condições de armazenamento (e.g. luminosidade e temperatura) e preparo (relação erva/água em massa, temperatura da água, número de infusões) da infusão de erva-mate afetam significativamente o conteúdo de polifenóis totais e conseqüentemente a atividade antioxidante dos extratos de erva-mate.

As metilxantinas são compostos nitrogenados heterocíclicos, dentre os quais os mais abundantes no reino vegetal são: a cafeína e a teobromina. Assim como os polifenóis, esse grupo de substâncias está naturalmente presentes em plantas utilizadas como base para alimentos e bebidas, como o café e a erva-mate (DE ARAGÃO et. al., 2009). A cafeína é conhecida por sua ação como estimulante do sistema nervoso central, sendo relacionada ao aumento da concentração e dos níveis de energia, assim como ao efeito termogênico das infusões de erva-mate (FIANI et al., 2021). Já a teobromina é considerada benéfica para a saúde cardiovascular, devido às propriedades vasodilatadoras que podem reduzir a pressão arterial e melhorar a circulação sanguínea (SHARIFI-ZAHABI et al., 2023).

As saponinas também são naturalmente encontradas na erva-mate. O nome deriva do fato desses compostos serem surfactantes naturais, com uma parte hidrofílica e uma lipofílica que gera a redução da tensão superficial da água, criando espuma quando agitado

(GOLEMANOV et al, 2014). A ingestão desses compostos está associada à ação anti-inflamatória e a redução do colesterol LDL no sangue, sendo benéfica para a saúde cardiovascular (GERBER et al., 2022; LUTOMSKI, GOZDZIEWSKA & LUSZCZKI, 2020).

Além desses grupos de compostos, a erva-mate possui em sua complexa composição, açúcares, vitaminas, minerais, lipídios, proteínas e celulose (CROGE, CUQUEL & PINTRO, 2022; WESTPHALEN et al., 2020). Devido à essa complexidade química, o emprego da análise instrumental associados à análise multivariada para o estudo da erva-mate tem se mostrado promissor (POLIDORO, et al., 2023, ROSSI, LOZANO & OLIVIERI, 2023).

2.5 FRAUDES EM ALIMENTOS

Além de ser uma matriz naturalmente complexa, a análise química de amostras alimentícias, com o intuito de identificar características associadas ao local de produção, pode ser comprometida por práticas fraudulentas. Segundo o Ministério da Agricultura, Pecuária e Abastecimento (MAPA, 2022), a fraude alimentar é uma prática desonesta caracterizada pela alteração proposital das características sensoriais, físicas, químicas ou biológicas de um produto de origem vegetal a fim de gerar benefício financeiro ou que possam prejudicar o consumidor. Essas práticas podem ser classificadas como:

Adulteração: redução, substituição, modificação total ou parcial da matéria-prima ou emprego de processos/substâncias não permitidas no tipo de produto;

Falsificação: imitação da forma ou rotulagem enganosa ou, emprego de denominação em desacordo com a classificação e a padronização correta do produto;

No Brasil, existe uma grande preocupação no que diz respeito à qualidade da erva-mate comercializada devido às adições, intencionais (fraude) ou não (sujidades), de materiais estranhos ao produto (BRAGHINI et al., 2014; SCHNEIDER et al., 2018). Estudos mostram que é corriqueiro encontrar misturas de materiais que alteram o sabor e a qualidade da erva-mate. Na maioria, são casos que relatam a presença de insetos e sujidades decorrentes do processo, mas em outros são identificadas adições de espécies vegetais de baixo custo ou talos para aumentar o peso do produto e de açúcar para mascarar sabores indesejados (HAURESKO et al., 2017; SCHNEIDER et al., 2018; SIGNOR et al., 2016).

Segundo a Resolução nº 716 da Agência Nacional de Vigilância Sanitária (ANVISA), caso a erva-mate seja produzida a partir de misturas com chás ou sejam feitas adições de açúcar ao produto, o rótulo deve apresentar essas informações para que as adições não sejam consideradas como adulterantes do produto (BRASIL, 2022). Entretanto, o uso de adulterantes

na indústria ervateira é impulsionada pela dificuldade de fiscalização e faz com que seja comum a comercialização de um produto de qualidade nutricional inferior, sem rastreabilidade de local de produção e sem garantia de segurança alimentar.

Vários estudos foram realizados a fim de identificar diferenças químicas em espécies vegetais de baixo custo que são utilizadas como adulterantes em amostras de erva-mate. Os resultados desses estudos indicam que as espécies adulterantes apresentam um perfil distinto de saponinas (PETROSELLI et al., 2018) e de compostos orgânicos voláteis (DALLAGO et al., 2011), além de apresentarem menor atividade antioxidante (FILIP et al., 2000), o que evidencia o fato de que a fraude não afeta apenas as características sensoriais dos produtos.

Apesar da literatura envolvendo a identificação da adição de adulterantes (e.g. talos ou açúcar) em amostras de erva-mate ainda ser escassa, estudos preliminares indicam a potencialidade de técnicas espectroscópicas e baseadas em imagens digitais sendo utilizadas com essa finalidade no estudo de matrizes alimentícias (DOGAN et al., 2023; VISCONTI et al., 2023; FERNANDES et al., 2023; YANG et al., 2023; CHAKRAVARTULA et al., 2022; KUCHARSKA-AMBROŹEJ & KARPINSK, 2020).

2.6 MÉTODOS ESPECTROSCÓPICOS

De maneira geral, métodos analíticos baseados em cromatografia ou espectroscopia são os mais empregados na quantificação de espécies químicas de interesse e na detecção de adulterações/contaminações em amostras alimentícias, tais como a erva-mate (SCHNEIDER et al., 2018, VIEIRA et al., 2020; PANZL et al., 2022; POLIDORO et al., 2023; PANZL et al., 2023; DE ALMEIDA et al., 2023; ERCIOGLU, VELIOGLU & BOYACI, 2018).

As técnicas espectroscópicas, tais como a espectroscopia na região do infravermelho (IV) e a espectroscopia de ressonância magnética nuclear (RMN), tem como vantagens o fato de, na maioria dos casos, serem rápidas, requererem mínimo preparo de amostras e de serem capazes de fornecer um grande volume de dados sobre o perfil químico das amostras. Entretanto, possuem como desvantagens a falta de sensibilidade para identificar quantidades traços e falta de seletividade na análise de matrizes complexas (KHARBACH et al., 2023).

2.6.1 Espectroscopia na região do infravermelho

A espectroscopia no infravermelho é fundamentada na análise das alterações que ocorrem no movimento das ligações presentes nos grupos funcionais específicos presentes na

molécula quando essa interage com a radiação, interações estas, principalmente relacionadas às vibrações moleculares fundamentais, como as vibrações de sobretons e combinações das ligações químicas. (FIGUEIRA, 2012). Moléculas que possuem ligações simétricas (e.g. H₂ e Cl₂) não sofrem variações no momento de dipolo durante a rotação ou vibração e por isso não podem ser analisadas por essa técnica (PAVIA et al., 2009; SKOOG, 2007).

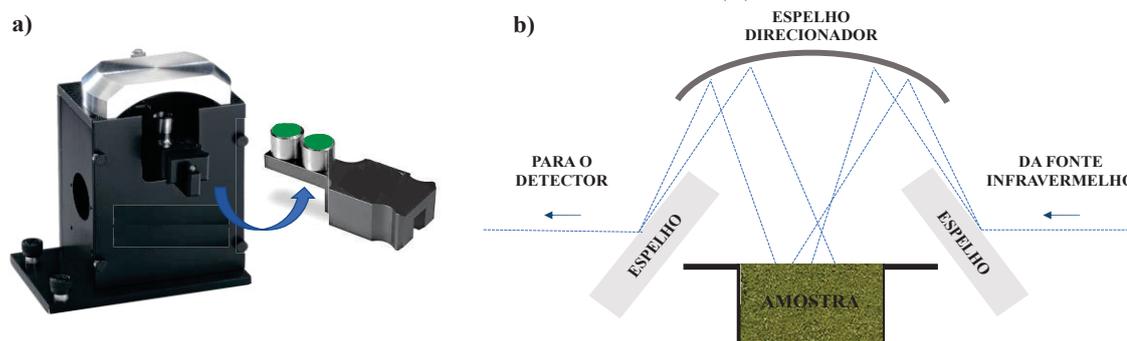
Essa região do espectro pode ser subdividida em três, que são denominadas de acordo com a sua distância em relação a região do visível: infravermelho próximo (NIR, do inglês *Near Infrared*); infravermelho médio (MIR, do inglês *Middle Infrared*) e infravermelho distante (FIR, do inglês *Far Infrared*) (SKOOG, 2007). Os espectros da região NIR são resultantes das combinações e sobreposições dos grupamentos O-H, N-H, S-H e C-H e fornecem informações sobre as interações intra e intermoleculares de grupos funcionais orgânicos. A região MIR é a mais utilizada atualmente e possui aplicações nas análises qualitativas e quantitativas de compostos orgânicos. Já a região FIR é utilizada principalmente na elucidação da estrutura de espécies inorgânicas e organometálicas (SILVERSTEIN, 2006).

A ampla abrangência da espectroscopia na região do infravermelho, em termos das características específicas de cada uma das regiões espectrais, oferece vantagens notáveis para o desenvolvimento de metodologias analíticas, sendo que a região ideal pode ser selecionada de acordo com as características da amostra e a finalidade da aplicação. Em estudos voltados à análise de matrizes alimentícias, a aplicação da técnica de infravermelho na região NIR tem se destacado no setor acadêmico e industrial, devido à rapidez e a potencialidade de miniaturização dos equipamentos (BEC, GRABSKA & HUCK, 2022; ZHANG et al., 2022).

Outra vantagem da técnica é a possibilidade de analisar amostras em praticamente todos os estados físicos, apenas fazendo a escolha do equipamento e acessórios adequados (MONTEIRO, 2011; SKOOG, 2007; PASQUINI, 2003). Com relação a amostras sólidas, um dos acessórios mais utilizados é o de refletância difusa com transformada de Fourier (DRIFT, do inglês *Diffuse Reflectance Infrared Fourier Transform*). Nesse tipo de acessório (FIGURA 5), um feixe da radiação é direcionado para a amostra através de um espelho côncavo e em seguida parte dessa radiação é refletida de forma difusa sendo direcionada ao detector por um espelho côncavo para a obtenção dos espectros (DIAS et al., 2016; CARVALHO, 2013).

Apesar das vantagens obtidas nas análises realizadas com o uso de acessórios como o DRIFT (e.g. rapidez, preparo mínimo de amostra), é necessário atentar-se para parâmetros que podem influenciar a qualidade dos espectros coletados, como, tamanho de partícula, teor de umidade, grau de compactação e ângulo de incidência, reduzindo a eficiência da obtenção de informações relevantes sobre o perfil químico das amostras (DE SOUZA & FERRÃO, 2006).

FIGURA 5 - REPRESENTAÇÃO DO ACESSÓRIO DRIFT ACOPLADO NA CÉLULA (A) E DO SEU FUNCIONAMENTO GERAL (B)



FONTE: A autora (2023).

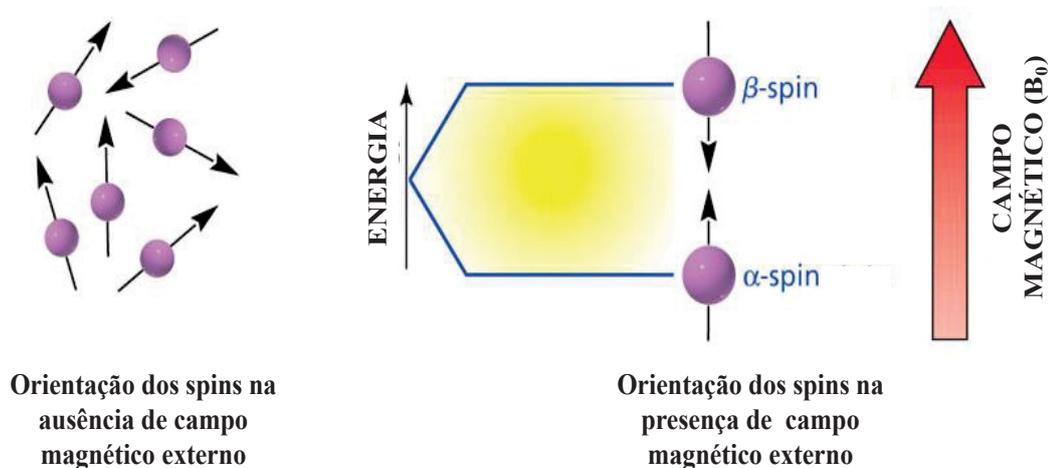
2.6.2 Espectroscopia de ressonância magnética nuclear

A espectroscopia de ressonância magnética nuclear (RMN) baseia-se nas propriedades dos núcleos atômicos que possuem um momento angular de spin e um momento magnético, como núcleos de hidrogênio (^1H) e carbono (^{13}C) (PAVIA et al., 2016). Esses núcleos se comportam como magnetos, e na ausência de um campo magnético seus spins nucleares se orientam de forma desordenada (FIGURA 6a). Na técnica de RMN, esses núcleos interagem com o campo magnético aplicado (B_0) permitindo a absorção de energia, fazendo com que entrem em ressonância magnética em frequências específicas, sendo promovidos de estados de spin de baixa energia (α -spin) para estados de spin de alta energia (β -spin) (FIGURA 6b) (BRUICE, 2007; TUGIZIMANA, PIATER & DUBERY, 2013).

Durante a etapa de relaxação para o estado fundamental, a energia recebida é dissipada e a emissão de sinais de ressonância induz a geração de corrente elétrica para o detector, resultando nos sinais químicos característicos da técnica de RMN em um espectro que registra a intensidade *versus* a frequência. Cada pico apresentado no espectro de RMN está associado ao tipo de núcleo atômico e à sua interação com o ambiente químico em que esse está inserido (PAVIA et al., 2016; BRUICE, 2007).

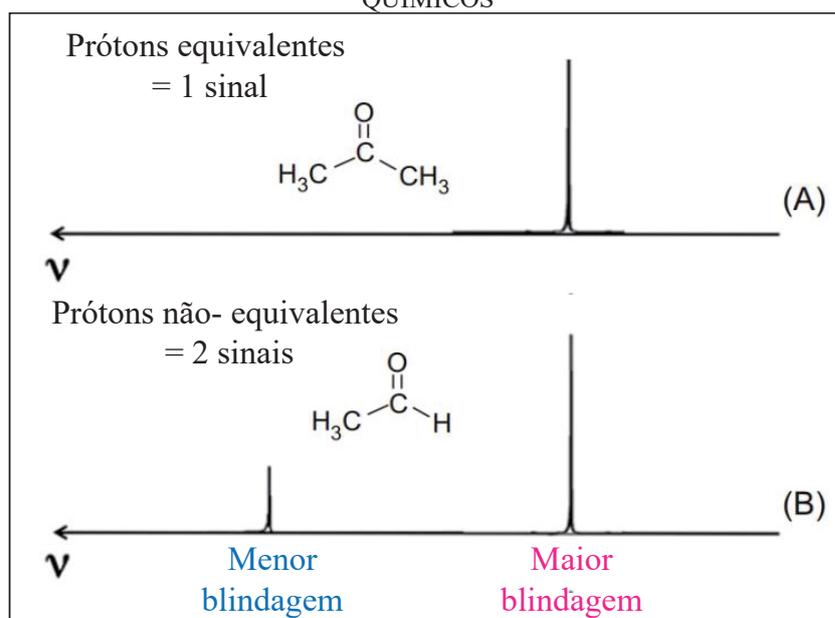
A nuvem eletrônica que cerca o núcleo possui o efeito de blindar parcialmente o campo magnético aplicado. Sendo assim, o campo magnético que efetivamente atinge o núcleo vai se diferenciar dependendo da densidade eletrônica do ambiente químico, tornando possível a diferenciação espectral entre os núcleos de acordo com seu deslocamento químico (NASCIMENTO, 2016). Em ambientes de alta densidade eletrônica, os núcleos são mais blindados e irão aparecer em frequências mais baixas no espectro (FIGURA 7). Já em ambientes de densidade eletrônica mais baixa, o campo magnético efetivo é maior e faz com que maiores frequências sejam necessárias para que os núcleos entrem em ressonância (BRUICE, 2007).

FIGURA 6 - EFEITO DA APLICAÇÃO DE UM CAMPO MAGNÉTICO NOS SPINS NUCLEARES



FONTE: Adaptado de Bruice (2007).

FIGURA 7 - EFEITO DA BLINDAGEM EM NÚCLEOS EXPOSTOS À DIFERENTES AMBIENTES QUÍMICOS



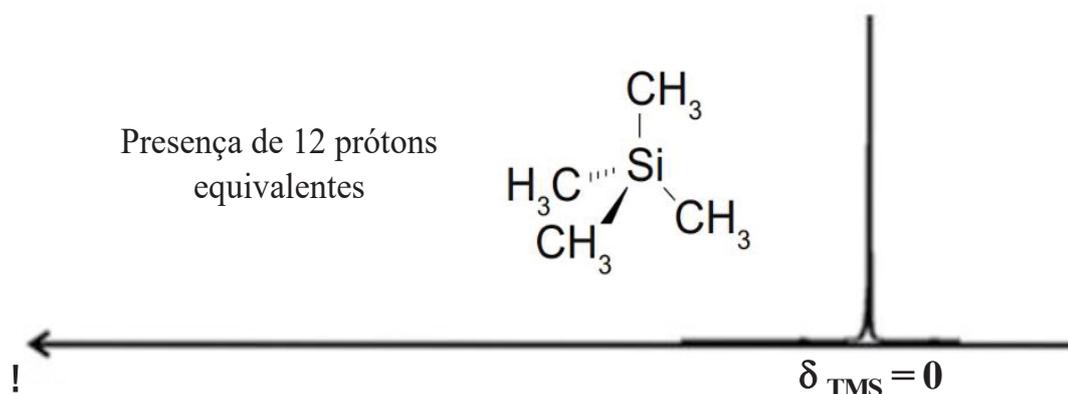
FONTE: Adaptado de NASCIMENTO, 2016.

Grupos de núcleos quimicamente equivalentes (sujeitos ao mesmo ambiente) irão gerar um sinal no espectro de RMN com um deslocamento químico característico (NASCIMENTO, 2016). O deslocamento químico (δ) pode ser expresso em partes por milhão (ppm) e representa a posição relativa dos sinais de ressonância dos núcleos presentes na amostra, quando esses são comparados com o padrão de referência (KELLER, 2010).

O tetrametilsilano (TMS) (FIGURA 8), é amplamente utilizado como referência no RMN devido à diversas propriedades, dentre elas: a) possuir todos os prótons em um mesmo

ambiente químico e altamente blindados, conseqüentemente gerando apenas 1 sinal espectral; b) ser quimicamente estável, o que reduz a chance de reações com a amostra ou com o solvente utilizado; c) apresentar sinal intenso devido ao número de prótons equivalentes; d) ser volátil, possibilitando a recuperação da amostra (NASCIMENTO, 2016).

FIGURA 8 – REPRESENTAÇÃO DO SINAL QUÍMICO DO TETRAMETILSILANO NO RMN

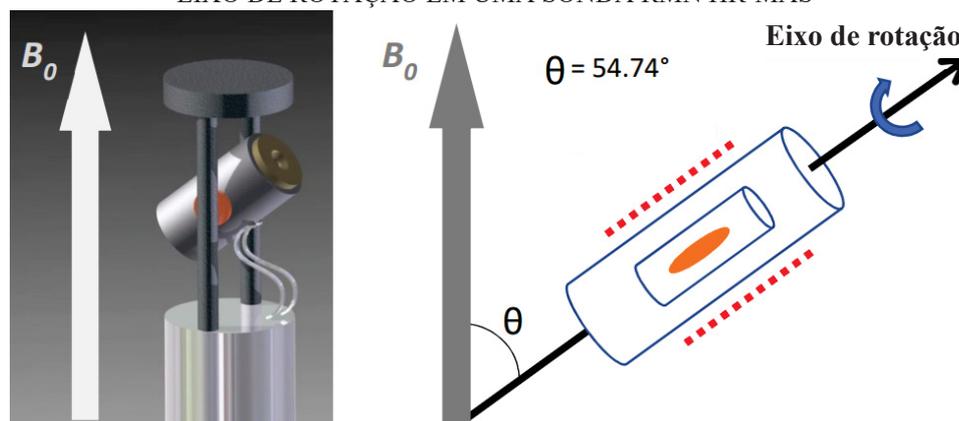


FONTE: A autora, 2023.

A espectroscopia de RMN pode ser aplicada a uma variedade de estados físicos de amostras, incluindo líquidos, emulsões e sólidos. Entretanto, amostras no estado sólido ou semissólido apresentam maior heterogeneidade e movimento molecular restrito, o que gera alargamento do sinal espectral (SANTOS et al., 2015). Esse alargamento é resultado das fortes interações dipolo-dipolo, presença de átomos com momentos magnéticos desemparelhados na amostra (aumento da susceptibilidade magnética) e do fato dos núcleos estarem inseridos em uma estrutura molecular que não possui simetria esférica (MAZZEI & PICCOLO, 2017).

O desenvolvimento de uma derivação da técnica, conhecida como RMN de Alta Resolução com Giro no Ângulo Mágico (HR-MAS, do inglês *High Resolution Magic Angle Spinning*), é capaz de aumentar a resolução dos espectros de amostras sólidas e semissólidas devido ao fato de que a rotação em ângulo mágico ajuda a reduzir as interações dipolo-dipolo e os efeitos de deslocamento químico que causam o alargamento (CORSARO et al., 2016). O “ângulo mágico” é um ângulo específico entre o eixo do rotor que contém a amostra e o campo magnético externo (FIGURA 9), que é definido em torno de $\theta=54,74^\circ$ para que o termo do componente $(3\cos^2\theta - 1)$ seja igual à zero, resultando em picos mais bem definidos (GOGIASHVILI et al., 2019; ALAM & JENKINS, 2012).

FIGURA 9 - REPRESENTAÇÃO DO ÂNGULO MÁGICO (θ) ENTRE O CAMPO MAGNÉTICO (B_0) E O EIXO DE ROTAÇÃO EM UMA SONDA RMN HR-MAS



FONTE: Adaptado de GOGIASHVILI et al., 2019.

A técnica HR-MAS possibilita a obtenção de informações estruturais valiosas de maneira rápida e não-destrutiva em amostras semissólidas, sendo particularmente importante na determinação do perfil químico de matrizes vegetais e alimentícias (SANTOS et al., 2023; AUGUSTIJN, GROOT & ALIA, 2021). Entretanto, heterogeneidade no tamanho das partículas da amostra e umidade elevada podem comprometer a qualidade do sinal obtido e, por isso é importante que as amostras estejam secas e maceradas antes de serem analisadas.

Outro ponto relevante em termos do preparo de amostras é a seleção do solvente que será colocado em contato com a amostra sólida para auxiliar na mobilidade dos compostos. Uma vez que, os núcleos presentes em solventes comuns (e.g. prótons provenientes da água ou metanol) podem mascarar os sinais dos núcleos presentes na amostra, é recomendado o uso de solventes deuterados (e.g. D_2O e CD_3OD) para contornar esse problema e também reduzir o potencial de interação entre o solvente e a amostra (NASCIMENTO, 2016).

Uma limitação da técnica de RMN na obtenção do perfil químico de amostras complexas é a potencial sobreposição de sinais, que compromete a identificação dos núcleos presentes. Nesses casos, a técnica de RMN bidimensional (RMN 2D) pode ser utilizada a fim de obter informações adicionais que permitam identificar com maior precisão os compostos presentes na amostra (GOGIASHVILI et al., 2019). Como o nome indica, os experimentos 2D possibilitam a coleta dos dados em duas dimensões, que são representadas em um mapa de contorno que correlaciona à intensidade de ressonância de determinadas frequências em dois momentos distintos (KEELER, 2002). Dentre os vários tipos de experimentos de RMN bidimensionais, se destacam os seguintes (KAISER, 2000; SZAKÁCS & SÁNTA, 2015):

COSY (CORrelation SpectroscopY): fornece informações sobre as correlações homonucleares entre prótons acoplados entre si, possibilitando a identificação de conexões covalentes entre eles e sendo útil para a determinação da multiplicidade dos sinais observados no espectro de RMN de ^1H .

HSQC (Heteronuclear Single Quantum Correlation): fornece informações sobre as correlações entre os sinais de núcleos de hidrogênio conectados com núcleos de carbono (ou outros heteroátomos) via ligações covalentes simples.

HMBC (Heteronuclear Multiple Bond Correlation): fornece informações sobre as correlações entre os sinais de núcleos de hidrogênio conectados com núcleos de carbono (ou outros heteroátomos) via ligação covalente dupla ou tripla.

2.7 MÉTODOS BASEADOS EM IMAGENS DIGITAIS

A contínua evolução de hardwares e softwares possibilitou o desenvolvimento de outra poderosa ferramenta analítica, a análise de imagens digitais (DAMASCENO et al., 2015). Esses métodos utilizam o processamento de imagens via análise multivariada para possibilitar a identificação e a extração de informações qualitativas e quantitativas sobre as amostras e vêm sendo amplamente aplicados em áreas, como na análise de alimentos (DINIZ, 2020), medicina (KOTINA, PLOSKIKH & SHIROKOLOBOV, 2022) e indústria (FERMO et al., 2021).

Segundo Pedrini & Schwartz (2008, p. 14), “Uma imagem pode ser definida como uma função de intensidade luminosa, denotada $f(x,y)$, cujo valor ou amplitude nas coordenadas espaciais (x,y) fornece a intensidade ou o brilho da imagem naquele ponto”. As informações que definem a cor, o brilho e outras propriedades visuais das imagens digitais estão contidas em um elemento da matriz bidimensional chamado *pixel* (do inglês, *Picture Element*) (SOLOMON & BRECKON, 2011).

A obtenção de uma imagem a partir dos valores numéricos que representam o nível de sinal em cada pixel se dá a partir do uso de modelo de cores (PEDRINI & SCHWARTZ, 2008). Existem vários modelos de cores que podem ser utilizados para representar imagens, cada um com características e aplicações específicas, sendo os mais difundidos: escala de cinza; RGB (sigla do inglês, *Red Green Blue*); HSI (sigla do inglês, *Hue Saturation Intensity*) e HSV (sigla do inglês, *Hue Saturation Value*) (GONZALEZ & WOODS, 2010).

Nas imagens em tons de cinza (FIGURA 10), cada pixel é representado por um valor numérico que indica o nível de intensidade luminosa naquele ponto da imagem. Esse número pode variar entre 0 (preto) e 255 (branco), sendo que os valores intermediários correspondem

aos diferentes tons de cinza (SOLOMON & BRECKON, 2011). A escala de cinza é normalmente utilizada em análises em que padrões de bordas, texturas e gradientes de intensidade são mais relevantes do que as informações de cor.

FIGURA 10 – MODELO DE IMAGEM DIGITAL REPRESENTADA EM ESCALA DE CINZA



FONTE: Adaptado de YE et al., 2022.

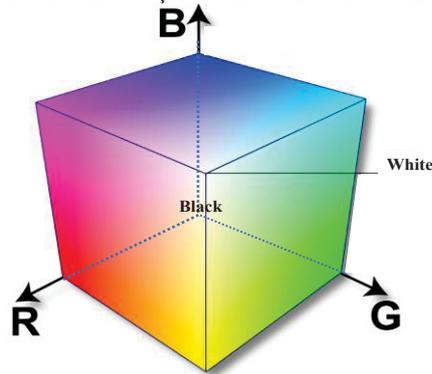
Em uma imagem colorida representada pelo modelo RGB (FIGURA 11), cada pixel é composto pela contribuição das cores vermelho, verde e azul em diferentes intensidades, resultando em uma ampla variedade de cores possíveis. A representação do modelo RGB se dá por meio de um espaço de cor tridimensional onde cada canal de cor possui valores que podem variar de 0 a 255 (SOLOMON & BRECKON, 2011). Como resultado da propriedade aditiva das cores que compõem o espaço RGB, a diagonal desse espaço tridimensional (FIGURA 12) configura a escala de cinza (ANNADURAI, 2023).

FIGURA 11 - REPRESENTAÇÃO DOS CANAIS DE COR DO MODELO RGB



FONTE: TAGLANG (2016).

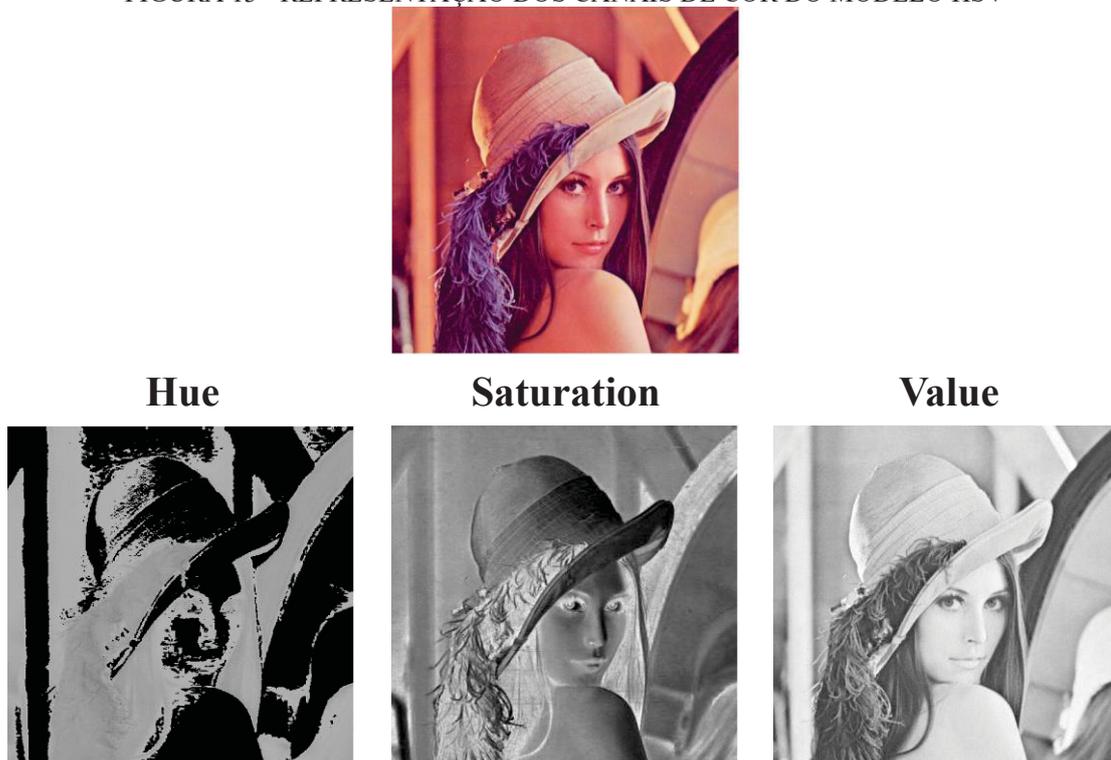
FIGURA 12 - REPRESENTAÇÃO DO ESPAÇO TRIDIMENSIONAL DO MODELO DE CORES RGB



FONTE: Adaptado de ANTONIADIS (2023).

Quando o objetivo da análise de imagens envolve a manipulação de brilho e luminosidade (FIGURA 13) ou a seleção mais precisa das cores, os modelos HSV e HSI apresentam resultados mais acurados do que o RGB. Esses modelos possuem semelhanças em suas estruturas e conceitos, sendo ambos aplicáveis na avaliação da tonalidade básica da cor (*Hue* ou matiz), e na determinação da pureza ou vivacidade da cor (*Saturation* ou saturação) (PEDRINI & SCHWARTZ, 2008). A principal diferença entre esses modelos é a de que enquanto no HSV o componente valor (V) representa a luminosidade percebida da cor, no HSI o componente intensidade (I) representa o brilho da cor (ANNADURAI, 2023).

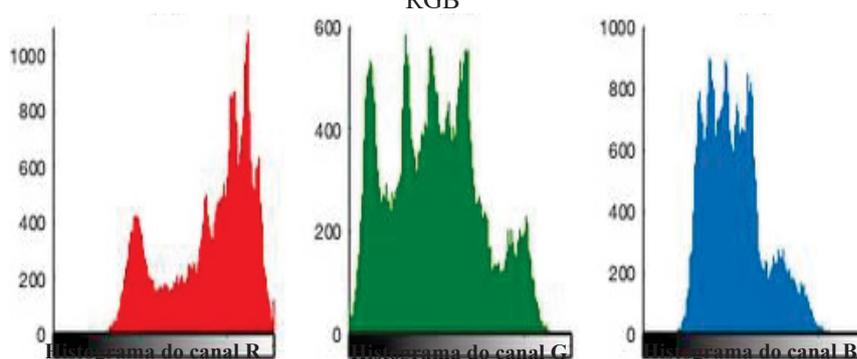
FIGURA 13 - REPRESENTAÇÃO DOS CANAIS DE COR DO MODELO HSV



FONTE: A autora (2023).

A distribuição dos pixels de uma imagem em cada modelo de cores pode ser avaliada em um histograma de cores. O histograma (FIGURA 14) é um gráfico que fornece a frequência relativa de ocorrência de cada componente do pixel possibilitando a obtenção de dados sobre a composição cromática da imagem, como a predominância de certas cores, a distribuição das intensidades ou luminosidades (SOLOMON & BRECKON, 2011).

FIGURA 14 - REPRESENTAÇÃO DOS HISTOGRAMAS DE COR OBTIDOS A PARTIR DO MODELO RGB



FONTE: Adaptado de REYES-LÓPEZ et al. (2021).

A análise de imagens digitais, seja a partir dos histogramas gerados por modelos de cores individuais ou combinados, vem sendo amplamente aplicada no desenvolvimento de metodologias analíticas, visto que a obtenção de dados fazendo uso de celulares, webcams ou scanners resulta em métodos rápidos, simples e de baixo custo para a análise de alimentos (DINIZ, 2020; DANTAS, 2015; GELADI et al., 1986).

2.8 QUIMIOMETRIA

Uma das principais características dos métodos instrumentais de análise e das metodologias baseadas em imagens digitais, é a obtenção de dados multivariados e a possibilidade do estudo simultâneo de variáveis, o que proporciona um entendimento mais abrangente sobre as amostras. Porém, juntamente com a quantidade de dados obtidos em cada análise, cresce a necessidade de ferramentas que possibilitem a extração e a interpretação das informações neles contidas, pois é comum que na presença de espectros muito complexos ou na existência de grandes semelhanças entre as amostras, a interpretação dos dados fique comprometida (CORREIA, 2007; FERREIRA, 2015).

A Quimiometria, uma subárea da Química Analítica, surgiu como resultado da busca por ferramentas matemáticas e estatísticas que fossem capazes de converter o sinal analítico

multivariado em informações úteis através de métodos apropriados (FERREIRA, 2015). Essas ferramentas podem ser subdivididas em três áreas principais: planejamento de experimentos, reconhecimento de padrões e calibração multivariada (GELADI, 2003):

Planejamento de experimentos – técnicas que permitem variar, de forma simultânea e sistemática, todas as variáveis envolvidas no processo de interesse utilizando uma quantidade reduzida de experimentos, de tal maneira que a influência de cada uma e a relação entre as mesmas seja avaliada para uma determinada resposta (BASTOS et al., 2007).

Reconhecimento de padrões – técnicas através das quais se torna possível encontrar relações que permitam a identificação de tendências, semelhanças e a classificação de amostras (OTTO, 2016; BRERETON, 2015; FERREIRA, 2015). São divididas em dois grupos: os métodos supervisionados, nos quais a classe da amostra é preestabelecida; e os não-supervisionados que não requerem informações prévias. Os métodos não-supervisionados mais utilizados são a Análise de Componentes Principais (PCA) e a Análise de Agrupamentos por Métodos Hierárquicos (HCA). No caso dos métodos supervisionados, destacam-se a Modelagem de Classes por Componentes Principais (SIMCA) e a Regressão por Mínimos Quadrados Parciais associada à Análise Discriminante (PLS-DA) (FERREIRA, 2015).

Calibração Multivariada – técnicas que consistem na construção de modelos matemáticos empíricos que permitam quantificar a correlação existente entre as propriedades de interesse (variável dependente) e o sinal analítico (variável independente). Os métodos mais utilizados são a Regressão pelo Método dos Mínimos Quadrados Parciais (PLS), a Regressão Linear Múltipla (MLR), Análise de Fatores Paralelos (PARAFAC), e a Resolução de Curvas Multivariadas (MCR) (FERREIRA, 2015).

2.8.1 Organização dos dados

A Quimiometria engloba um conjunto de técnicas onde os dados são processados de forma a fornecerem informações de natureza química. Para isso é necessária a organização dos dados em matrizes adequadas ao tratamento quimiométrico (BRERETON, 2015). A matriz de dados (\mathbf{X}) será constituída por m linhas (amostras) e n colunas (variáveis), sendo que as variáveis podem ser tanto quantitativas quanto qualitativas.

Após a organização é necessário efetuar a observação dos dados originais (brutos). Essa etapa é crucial visto que em alguns casos, amostras com comportamento anômalo podem ser detectadas visualmente, devendo ser reavaliadas para não comprometerem a qualidade do tratamento quimiométrico (BEEBE, 1998; FERNANDES, 2006).

2.8.2 Pré-processamento dos dados

Em muitos casos, os dados brutos apresentam variações aleatórias ou sistemáticas que podem influenciar os resultados das análises multivariadas (BEEBE, 1998). Nesse contexto, a aplicação de técnicas de pré-tratamentos para reduzir as variações indesejáveis se torna extremamente importante para garantir a qualidade dos resultados obtidos nas análises posteriores (e.g. PCA, PLS, PLS-DA). Esses pré-tratamentos podem ser divididos em duas classes: quando aplicados às amostras são chamados de transformação e quando aplicados às variáveis são chamados de pré-processamento (FERREIRA, 2015).

Vale ressaltar que as técnicas de pré-tratamento causam modificações nos dados brutos e se aplicadas de maneira incorreta podem levar à perda de informações relevantes. A escolha do pré-tratamento adequado depende do conhecimento sobre as amostras e sobre os processos químicos e físicos envolvidos na análise (BRERETON, 2015).

De forma geral, os métodos de centragem dos dados na média, correção de linha base, alisamento e escalamento são os mais utilizados na literatura, e além disso, técnicas de derivação, normalização, correção de espalhamento de sinal também vêm sendo empregadas (LARA et al., 2017; RIBEIRO et al., 2017; SOARES et al., 2017; SOUZA et al., 2018).

No processamento de dados espectrais provenientes de amostras sólidas, os processamentos mais aplicados são as técnicas de alisamento, correção de espalhamento multiplicativo (MSC), centragem dos dados na média e alinhamento de sinais. Para aplicações em dados com variáveis que apresentam ordem de grandeza significativamente distintas entre si é indicado o uso do autoescalamento (FERREIRA, 2015).

2.8.2.1 Centragem dos dados na média

A centragem dos dados na média é baseada na translação dos eixos para um valor médio, resultando no deslocamento dos dados para a origem do sistema de coordenadas, mas garantindo que a estrutura dos mesmos seja preservada. Inicialmente a média de cada coluna de dados é calculada e em seguida essa média é subtraída do valor de cada uma das variáveis da coluna (FERREIRA, 2015; MARTINS, 2015).

2.8.2.2 Escalamento

O autoescalamento consiste na obtenção do valor centrado na média da forma descrita anteriormente, seguida da divisão desse valor pelo desvio padrão da coluna. O escalamento de

Pareto é muito semelhante ao autoescalamamento (BERG et al., 2006). No entanto, em vez do desvio padrão, é utilizado o valor da raiz quadrada do desvio padrão como fator de divisão. Nos dois casos a estrutura dos dados é completamente preservada e as diferenças e variações existentes podem ser melhor visualizados (FERREIRA, 2015; BEEBE, 1998).

2.8.2.3 Correção multiplicativa de espalhamento e padronização normal de sinal

As técnicas de MSC (do inglês, *Multiplicative Scatter Correction*) e SNV (do inglês, *Standard Normal Variate*) são comumente aplicadas em dados espectroscópicos para corrigir diferenças de espalhamento causados por fenômenos físicos, como alterações no caminho óptico, flutuações de energia e temperatura, assim como pela falta de homogeneidade no tamanho e na forma de partículas sólidas; (SOUZA et al., 2012; FERREIRA, 2015). Apesar de serem obtidos por modelos matemáticos de autoescalamamento distintos, o comportamento espectral dos dados tratados pelos dois pré-processamentos é bastante semelhante (ALMEIDA, 2009).

2.8.2.4 Técnicas de alisamento (suavização)

As técnicas de alisamento, atuam na redução do ruído aleatório existente nos dados. Os métodos são baseados no aumento da razão sinal-ruído, levando à separação entre o sinal “real” e o ruído (FERREIRA, 2015; BEEBE, 1998). Os métodos mais comuns atuam como filtros lineares que envolvem a definição de um segmento (“janela”) de pontos que se desloca pelo espectro reduzindo todos os pontos à um ponto central (BRERETON, 2015). Existem diversas variações da técnica (e.g. método da média e alisamento com filtro de Fourier), mas o mais utilizado é o de suavização de Savitzky-Golay (FERREIRA, 2015).

2.8.2.5 Alinhamento de sinais espectrais

Os métodos de alinhamento têm o objetivo de minimizar os efeitos das variações nas posições de sinais espectrais causadas por efeitos físico-químicos nas amostras. O método mais utilizado em dados de RMN é o *bucketing*, no qual o espectro é dividido em regiões não sobrepostas com largura previamente definida para gerar um novo conjunto de dados que abrange a variabilidade do sinal (SOUZA et al., 2013). Quando os desalinhamentos não são corrigidos pelo *bucketing*, são utilizados outros métodos de alinhamento, como, o Alinhamento por Correlação Otimizada (COW), o Deslocamento correlacionado (*coshift*) e o Deslocamento correlacionado por intervalos (*Icoshift* do inglês, *Interval correlation-shifting* (SOUZA, 2020).

2.8.3 Análise de componentes principais (PCA)

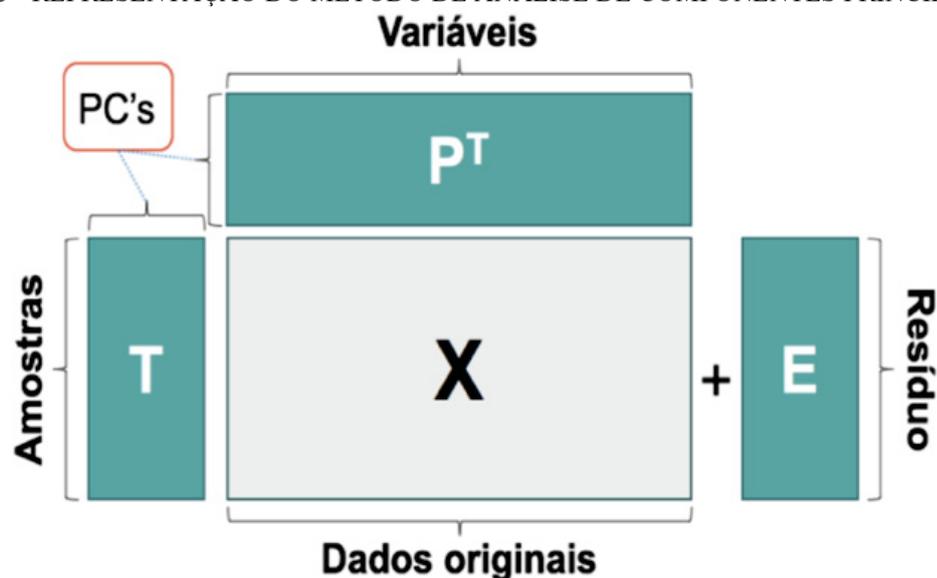
A PCA (do inglês, *Principal Component Analysis*) é a técnica de reconhecimento de padrões não-supervisionada mais utilizada, sendo considerada a base da análise multivariada (CHATFIELD, 2013). Seu objetivo principal é reduzir a dimensionalidade dos dados através da transformação das variáveis originais em suas componentes principais, que são um conjunto de vetores ortogonais não-correlacionadas capazes de explicar a variância existente nos dados (JOLLIFE, 1986; CHATFIELD, 2013). Cada componente principal (PC) possui um conjunto de autovalores e autovetores (PARREIRA, 2003).

- **Autovalores:** São os valores que representam a contribuição relativa de cada PC na explicação da variação total dos dados. O tamanho deste valor é uma indicação direta da importância da PC para a explicação da variância existente nos dados.
- **Autovetores:** São valores que representam o peso de cada variável original sobre cada PC. Os autovetores são escalados como coeficientes de correlação e variam de +1,0 a -1,0. Quanto mais próximo desses extremos se encontra o autovetor, maior é a importância da variável para a componente principal em questão.

Os resultados da PCA podem ser interpretados como a projeção dos dados originais em um espaço dimensional reduzido capaz de representar os padrões existentes entre amostras. Essa “compressão dos dados” resulta da capacidade da PCA de agrupar as variáveis que fornecem informações semelhantes e permite não só a comparação entre as amostras de acordo com os agrupamentos, mas também a detecção de amostras com comportamento anômalo que podem interferir na interpretação dos resultados (FERREIRA, 2015).

Matematicamente, o método da PCA pode ser descrito como a decomposição da matriz de dados original \mathbf{X} em duas novas matrizes (FIGURA 15): a matriz de scores (\mathbf{T}) cujos dados representam as coordenadas das amostras em cada PC, podendo ser visualizados através de gráficos bidimensionais ou tridimensionais; a matriz de *loadings* (\mathbf{P}^T) que representa o peso de cada variável original na formação dos novos eixos; além disso, temos a representação dos resíduos (\mathbf{E}) visto que, devido a existência de ruído instrumental, não é possível explicar 100% da variância dos dados (FERRÃO et al., 2007).

FIGURA 15 - REPRESENTAÇÃO DO MÉTODO DE ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)



FONTE: A autora (2023).

2.8.5 Regressão pelo método de mínimos quadrados parciais (PLS)

A regressão pelo método de PLS (do inglês, *Partial Least Squares*) se baseia no método de PCA, porém o objetivo do método é a construção de um modelo capaz de explicar a relação existente entre as variáveis independentes (X) e as variáveis dependentes (y). O nome dado à essa técnica está relacionado com o fato de que cada parâmetro do modelo criado será estimado iterativamente como a inclinação de uma regressão simples pelo método dos mínimos quadrados e o termo “parciais” provém do fato do vetor ser considerado fixo a nova cada fase de iteração (WOLD et al., 2001).

De modo similar ao que ocorre na PCA, as matrizes de dados também são decompostas em uma matriz de *scores* e uma de *loadings* e em seguida uma relação linear multivariada é estabelecida entre elas. A decomposição matricial gera as variáveis latentes (LV do inglês, *Latent Variables*), que são combinações lineares das variáveis originais geradas de modo a maximizar a covariância entre X e y (GOMES, 2013).

O desenvolvimento de métodos de calibração multivariada consiste em três etapas principais, sendo elas, a calibração, a validação e a predição. Durante a etapa de **calibração**, os dados da matriz X (e.g. espectros NIR) são correlacionados com o vetor y (e.g. concentração de açúcar) por meio da técnica de regressão adequada para a geração do modelo de regressão. Em seguida, na etapa de **validação**, um novo grupo de amostras com concentrações conhecidas é analisado a fim de avaliar a qualidade do modelo matemático construído. A validação pode ser interna (utilizando amostras que participaram da calibração) ou externa (utilizando um

conjunto distinto de amostras). Após a otimização e avaliação da qualidade do modelo, esse pode ser aplicado a um conjunto de amostras com concentrações desconhecidas para que sua capacidade de **predição** será avaliada (FERREIRA, 2015; FERRÃO et al., 2007).

A seleção das amostras que farão parte dos conjuntos de calibração e validação é uma etapa crucial para garantir que o modelo seja representativo da população e que seja capaz de resultar em predições precisas. Usualmente, 80% das amostras do conjunto são utilizadas na calibração e 20% na validação, e a seleção se dá por meio de métodos de seleção de amostras.

O algoritmo de *Kennard-Stone* (KENNARD & STONE, 1969), que se baseia na distância euclidiana entre as amostras, está entre os métodos de seleção mais difundidos em análises quimiométricas. O algoritmo inicia identificando as duas amostras mais distantes, visando garantir que a amplitude natural existente seja representada no modelo. Nas iterações subsequentes, as demais amostras são selecionadas de maneira que se obtenha um subconjunto de amostras que represente o máximo da variabilidade do conjunto original.

A otimização de modelos PLS consiste na determinação do número ideal de variáveis latentes (VL), visto que a seleção de muitas VL's aumenta o risco de sobreajuste e de introdução de ruído ao modelo, enquanto que o uso de um número de VL menor que o ideal pode causar perda de informações relevantes. (DESPAGNE & MASSART, 1998).

A seleção do número ideal de VL pode ser realizada através de métodos de validação cruzada, sendo que o mais comumente aplicado é o “*leave one out*”. Neste método uma das amostras do conjunto de calibração é excluída da fase de desenvolvimento do modelo, sendo reservada como elemento de predição. Este processo é repetido n vezes, de maneira a permitir que todas as amostras participem como elementos de predição (FERREIRA et al, 1999).

Outra etapa da otimização dos modelos de calibração multivariada, é a utilização dos valores de Importância da Variável na Projeção (VIP, do inglês *Variable Importance in Projection*). Essa ferramenta possibilita a identificação das variáveis que contribuem de forma significativa para o estabelecimento das correlações criadas pelo modelo PLS, sendo que quanto maior o valor de VIP, maior a importância da variável no modelo e, de maneira geral, variáveis com valores de $VIP > 1$ são consideradas relevantes (TSAGKARIS et al., 2023; BELCHIOR, BOTELHO & FRANCA, 2022). A identificação do nível de importância das variáveis permite então o descarte de variáveis irrelevantes, simplificando a interpretação dos modelos.

A avaliação da qualidade dos modelos de calibração multivariada pode ser efetuada com base em diversas figuras de mérito, dentre as quais se destacam: a avaliação dos erros quadráticos médios de calibração (RMSEC, do inglês *Root Mean Square Error of Calibration*), validação cruzada (RMSECV, do inglês *Root Mean Square Error of Cross-validation*) e de

predição (RMSEP, do inglês *Root Mean Square Error of Prediction*); e a análise dos coeficientes de determinação de validação cruzada (R^2_{CV}) e de predição (R^2_P). Os parâmetros RMSE e R^2 podem ser calculados pelas equações (1) e (2):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (1)$$

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

Onde:

\hat{y} = valores preditos;

y = valores mensurados;

\bar{y} = valores médios;

n = número de amostras.

A qualidade de modelos multivariados pode ser afetada pela presença de amostras com comportamento anômalo (*outliers*), ou seja, que apresentem um perfil de resposta diferente das demais. Em muitos casos, essas amostras precisam ser removidas do conjunto amostral, pois são resultado de erros de análise que podem gerar predições com RMSE mais elevados (BRERETON, 2015; FERREIRA, 2015).

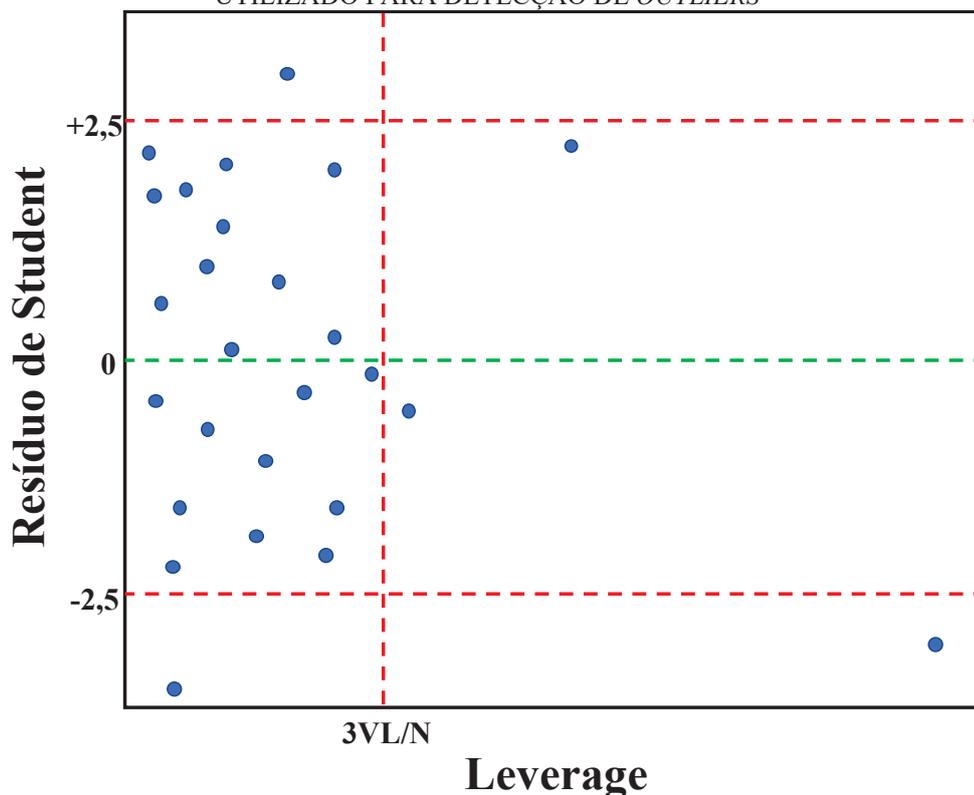
No entanto, se após a reavaliação criteriosa, a opção de erros sistemáticos puder ser descartada, é importante que as amostras com comportamento diferenciado sejam mantidas devido ao potencial de conterem informações relevantes ao modelo. Nesse caso, idealmente novas amostras que apresentem o mesmo perfil de anomalia devem ser incluídas na modelagem para que o ajuste seja aprimorado ((BRO & SMILDE, 2014; ABDI & WILLIAMS, 2010).

A detecção de amostras com comportamento anômalo pode ser efetuada a partir de um gráfico (FIGURA 16) que correlaciona dois parâmetros estatísticos, a influência (*leverage*) e o resíduo de *Student* (OLIVEIRA, 2018; FERREIRA, 2015).

O valor de *leverage* indica a capacidade de uma amostra em influenciar significativamente o modelo, sendo que altos valores nesse parâmetro indicam que o perfil da amostra difere do perfil médio do restante do conjunto. O limite que determina quais amostras devem ser avaliadas em relação à sua influência (linha vertical na FIGURA 16) é calculado a partir do número de variáveis latentes (VL) e do número de amostras existentes no conjunto de calibração (n) através da equação (3) (FERREIRA, 2015; OLIVEIRA, 2018):

$$Leverage = \frac{3VL}{n} \quad (3)$$

FIGURA 16 - REPRESENTAÇÃO DO GRÁFICO DE *LEVERAGE VS* RESÍDUOS DE *STUDENT* UTILIZADO PARA DETECÇÃO DE *OUTLIERS*



FONTE: A autora (2023).

Os resíduos de *Student* indicam a presença de amostras que não foram modeladas adequadamente pelo modelo e que por isso, apresentam uma diferença significativa entre os valores reais e os preditos pelo modelo multivariado. Os valores dos resíduos estudentizados são calculados dividindo o resíduo pelo erro padrão ajustado. Como diretriz geral, amostras com resíduos maiores que +2,5 ou menores que -2,5 (linhas horizontais vermelhas na FIGURA 16) devem ser reavaliadas e se necessário, retiradas dos modelos de calibração para que não afetem a capacidade de predição do mesmo (LU et al., 2023). Entretanto, o valor dos limites pode ser ajustado de acordo com a natureza dos dados e o número de amostras no modelo.

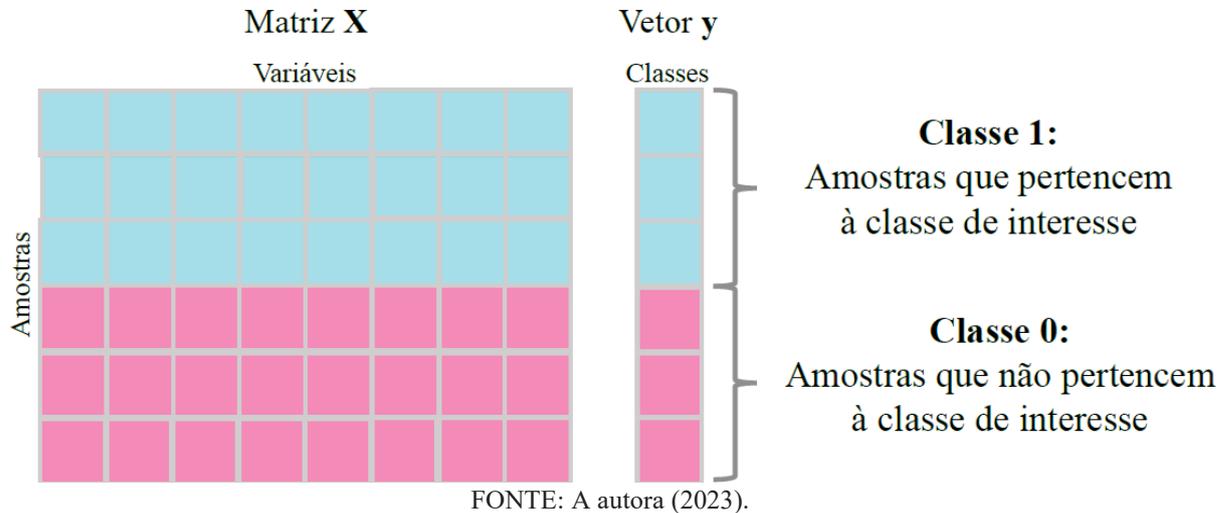
2.8.6 Análise discriminante pelo método de mínimos quadrados parciais (PLS-DA)

O PLS-DA (do inglês, *Partial Least Square-Discriminant Analysis*) utiliza a mesma estrutura de construção de modelos que o PLS. No entanto, o vetor y corresponde a variáveis categóricas que descrevem a atribuição de classe das amostras, sendo convencional, atribuir o valor 1 para a classe de interesse e 0 para a outra classe (FIGURA 17).

Devido ao fato de as variáveis categóricas terem sido modeladas de maneira contínua, os valores de classe previstos pelo modelo geralmente não serão iguais a 0 e 1. Por esse motivo,

é necessário o estabelecimento de um valor limite (*threshold*) que permita definir objetivamente se a classe da amostra foi ou não predita corretamente (SANTANA et al., 2020).

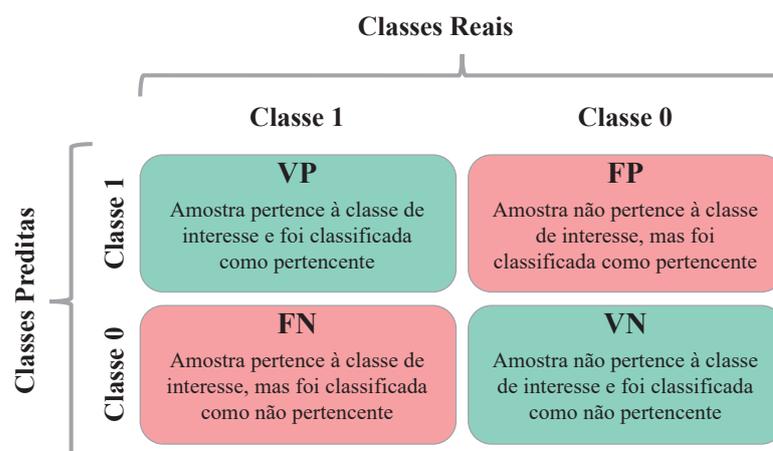
FIGURA 17 - REPRESENTAÇÃO DA MATRIZ X E DO VETOR Y UTILIZADOS NO MÉTODO PLS-DA



O *threshold* pode ser fixo ou variável (baseado no teorema de Bayes). Se for fixo, apenas amostras com valores preditos $> 0,5$ serão classificados como pertencentes à classe de interesse, enquanto que no *threshold* variável a amostra será designada à classe na qual tiver a maior probabilidade de pertencer. A maior diferença entre as abordagens é a de que no método variável todas as amostras serão alocadas em uma classe, enquanto no fixo existe a possibilidade de amostras não serem alocadas em nenhuma das classes (WESTAD, 2015).

O desempenho dos modelos de classificação pode ser avaliado com base na interpretação da matriz de confusão. A matriz de confusão (FIGURA 18) é uma tabela que mostra as frequências de classificação do modelo, resumizando o número de predições corretas e incorretas obtidas pelo mesmo. Os resultados apresentados na matriz de confusão permitem caracterizar os erros em relação a taxa de falsos positivos (FP), falsos negativos (FN), verdadeiros positivos (VP) e verdadeiros negativos (VN) e obter os valores para métricas de desempenho, tais como, sensibilidade, especificidade e precisão (KRSTINIĆ et al., 2020).

FIGURA 18 - REPRESENTAÇÃO DA MATRIZ DE CONFUSÃO E CARACTERIZAÇÃO DOS TIPOS DE ERROS DE CLASSIFICAÇÃO



FONTE: A autora (2023).

A avaliação das métricas, sensibilidade, especificidade e precisão possibilita a avaliação do desempenho dos modelos de classificação. A **sensibilidade** indica a capacidade do modelo em classificar corretamente as amostras pertencentes a uma classe (VP). A **especificidade** avalia a capacidade do modelo em identificar corretamente as amostras que não pertencem a uma classe (VN). Por fim, a **precisão** indica a proporção de classificações corretas, independente da classe da amostra (BAJOUB et al., 2017). Essas métricas são calculadas a partir das equações (4), (5) e (6), respectivamente:

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \quad (4)$$

$$\text{Especificidade} = \frac{VN}{VN + FP} \quad (5)$$

$$\text{Precisão} = \frac{VN + VP}{VN + VP + FN + FP} \quad (6)$$

2.9 APLICAÇÕES DE QUIMIOMETRIA NA ANÁLISE DE ALIMENTOS

A potencialidade das técnicas de reconhecimento de padrões, classificação e calibração multivariada vêm sendo exploradas em combinação com as técnicas espectroscópicas e de análise de imagens no estudo de matrizes alimentícias principalmente para a classificação de amostras de acordo com a origem, espécie ou marca (AZEVEDO et al., 2019; KOCH et al., 2018; MARCELO et al., 2014), para o acompanhamento do perfil químico (FERNANDES et al., 2017; QI et al., 2018) e para a determinação de propriedades de interesse (KONIECZYNSKI et al., 2017; DE LUCA et al., 2018; LI et al., 2018; MAZUR et al., 2014).

Com foco na discriminação de amostras arroz de acordo com a origem geográfica e técnicas de cultivo, Wu e colaboradores (2023) utilizaram a PCA e a técnica de PLS-DA em dados espectrais NIR. Com base na PCA, os autores inicialmente observaram grande sobreposição das amostras, sem uma clara distinção entre as provenientes do cultivo orgânico e convencional e sem a formação de agrupamentos relacionados à origem geográfica. Com o intuito de remover possíveis interferências e melhorar a discriminação entre as amostras, os autores aplicaram métodos de suavização e de correção de espalhamento multiplicativo aos espectros. Como resultado, a PCA indicou a tendência de agrupamento entre as técnicas de cultivo e o modelo de PLS-DA resultou em uma precisão de 89,7%. Com relação à origem geográfica, o modelo PLS-DA resultou em uma precisão de 100% para as amostras do Nordeste da China e acima de 90% para as demais regiões avaliadas.

A espectroscopia na região do infravermelho associada às técnicas de reconhecimento de padrões (HCA e PCA) também foi aplicada por Schneider et al. (2018) no desenvolvimento de uma metodologia para detecção de adulterações de amostras de erva-mate pela adição de sacarose. As soluções de sacarose na faixa entre 5-30% foram misturadas às amostras não-adoçadas de erva e em seguida os espectros foram obtidos. A região de impressão digital (*fingerprint*) foi selecionada e vários métodos de pré-processamento foram aplicados aos dados. O dendrograma resultante da HCA mostrou a formação de um cluster onde foram agrupadas as amostras sem adição de açúcar. Apesar de existir a tendência de agrupamento das amostras com maiores teores de açúcar, a HCA não levou à formação de clusters de acordo com a concentração de sacarose. Na PCA, a primeira PC, responsável por 79% da variância dos dados, possibilitou a visualização de agrupamentos de amostras com e sem adição de açúcar, mas assim como a HCA, não houveram agrupamentos de acordo com a concentração.

Truzzi et al. (2023), avaliaram o potencial da técnica de espectroscopia de RMN de ^1H associada ao PLS-DA para o controle de qualidade de chocolate amargo, com foco na identificação e quantificação da presença de substituintes da manteiga de cacau acima do limite permitido pela legislação. Para isso, amostras de chocolate sem gordura foram misturados com um mix de gorduras substituintes (e.g. palma e manga) na faixa de 0 a 50%. Os espectros de RMN de ^1H das misturas e das gorduras puras foram utilizados na construção do banco de dados no qual foram selecionadas as regiões correspondentes à região de *fingerprint* dos glicerídeos. Modelos PLS-DA foram construídos para a identificação do tipo de substituinte adicionado à mistura e apresentação resultados de sensibilidade entre 60-100%. O método de regressão via PLS foi aplicado na quantificação dos substituintes e resultou em modelos com R^2 de predição de aproximadamente 0,97, enquanto os valores de RMSE variaram na faixa de 1,5-5,7%.

A espectroscopia de RMN de ^1H também foi utilizada por Santos, Alvarega e Boffo (2020) no desenvolvimento de uma metodologia para discriminação de amostras de café de acordo com a qualidade (tradicional, superior ou gourmet). As regiões espectrais que possuíam sinais interferentes foram removidas e os espectros foram pré-processados com o uso do método de métodos de autoescalamento e alinhamento de picos. Através da PCA foram identificados sinais característicos que possibilitaram o agrupamento das amostras. Foi possível identificar que as amostras provenientes do estado da Bahia formaram um agrupamento por apresentarem os menores teores de quinina quando comparadas as amostras de outras regiões. Um agrupamento caracterizado pelos maiores teores de lipídeos foi formado por amostras da região da Chapada Diamantina. Além disso, os cafés de maior valor comercial (superior/gourmet) apresentarem os maiores teores de cafeína, trigonelina e ácidos cafeoilquínicos.

Pérez e Botteron (2013), implementaram um método baseado no processamento de imagens digitais no modelo de cores HSL para a determinação do teor de galhos presente em amostras de erva-mate. As amostras foram misturadas com porcentagens de galhos variando entre 15-45% e em seguida, os histogramas gerados a partir das imagens obtidas em um scanner comum foram analisados a partir de duas técnicas: uma baseada na área calculada através dos pixels da imagem e outro baseado no volume. Tanto a área quanto o volume foram calculados a partir de equações que estabelecem a correlação entre a proporção de galhos e folhas existente na amostra. Os resultados indicaram que, o método foi capaz de estimar a porcentagem de galhos presente nas amostras de erva com um erro de aproximadamente 4% e coeficiente de determinação (R^2) de 0,9997, mostrando que essa poderia ser uma alternativa para os métodos usados tradicionalmente.

A análise de imagens digitais associada à métodos supervisionados foi aplicada por Milanez & Pontes (2015) na verificação de adulterações por adição de óleo de soja em amostras de azeite de oliva extra virgem (AOEV) e na discriminação das amostras em relação à marca. Para esse fim, o histograma do espaço de cor RGB foi utilizado para construção de modelos de baseados na LDA com seleção de variáveis pelo algoritmo de projeções sucessivas (SPA) na PLS-DA. Os resultados de classificação em relação à marca, obteve uma acurácia de 100% para ambos os modelos (LDA e PLS-DA). O estudo envolvendo a detecção de adulteração com óleo de soja em amostras de AOEV foi realizado separadamente para cada marca, sendo que as misturas foram preparadas com até 10% de óleo de soja. A taxa de classificação correta para as amostras de predição da marca A foram de 88% e 94% pelos modelos SPA-LDA e PLS-DA, respectivamente. No das marcas B e C, todas as amostras de previsão foram classificadas corretamente por ambos os modelos.

Apesar das inúmeras vantagens do uso de metodologias quimiométricas (e.g. rapidez, baixo custo, preparo de amostra simplificado) e do crescimento do número de aplicações desenvolvidos com foco no estudo de matrizes alimentícias, a literatura disponível sobre aplicações desse tipo de metodologia a amostras de erva-mate ainda é escassa, principalmente com foco na discriminação das amostras de acordo com uma origem geográfica específica ou na avaliação da presença de adulterantes no produto. Sendo assim, esse trabalho propõe a avaliação da potencialidade dos métodos quimiométricos para extrair informações químicas relevantes a partir dos dados espectrais e de imagens digitais de amostras de erva-mate.

3 OBJETIVOS

3.1 OBJETIVO GERAL

Desenvolvimento de estratégias analíticas baseadas na associação entre técnicas espectroscópicas, imagens digitais e análise multivariada que possibilitem a discriminação de amostras de erva-mate de acordo com a origem geográfica, assim como a determinação da presença de adulterantes em amostras comerciais.

3.2 OBJETIVOS ESPECÍFICOS

- a) Analisar o perfil químico de amostras comerciais de erva-mate (*Ilex paraguariensis* st. *Hill*) utilizando a técnica de Ressonância Magnética Nuclear HR-MAS de ^1H aliada a metodologias quimiométricas. E, em seguida, aplicar a técnica de RMN bidimensional para a elucidação estrutural dos principais marcadores;
- b) Investigar o perfil químico de amostras comerciais de erva-mate (*Ilex paraguariensis* st. *Hill*) utilizando a técnica de espectroscopia na região do infravermelho (NIR e MIR) aliada a metodologias quimiométricas para a avaliação de similaridades e distinções entre amostras de erva-mate provenientes de diversas regiões e para a identificação de amostras contendo sacarose;
- c) Otimizar os parâmetros de aquisição de imagens digitais a fim de desenvolver um modelo de regressão PLS com base nos histogramas de cores para a determinação quantidade de talos presentes em amostras de erva-mate.

4 MATERIAL E MÉTODOS

4.1 AMOSTRAS

Durante esse estudo, foi analisado um total de 130 amostras de erva-mate adquiridas em estabelecimentos comerciais de quatro estados brasileiros: Mato Grosso do Sul (2), Paraná (37, sendo 22 com IP), Santa Catarina (24), Rio Grande do Sul (68, sendo 10 adoçadas). Entretanto, devido à alta degradabilidade das amostras e às particularidades de cada metodologia, os conjuntos amostrais de cada uma das etapas experimentais possuem um número particular de amostras. As principais informações fornecidas nas embalagens das amostras, incluindo a origem geográfica declarada pelos produtores, o tipo de produto e o tipo de embalagem, são apresentadas na TABELA 1.

TABELA 1 - Identificação das amostras de erva-mate e principais informações fornecidas nos rótulos (continua)

Código	Origem	Estado	Embalagem	Tipo	Etapas
A01	São Mateus (IG)	PR	Vácuo Laminada (VL)	Nativa	MIR/NIR (PCA)
A02	São Mateus (IG)	PR	Papel (PP)	Exportação	MIR/NIR (PCA)
A03	São Mateus (IG)	PR	Papel (PP)	Tradicional	MIR/NIR (PCA)
A04	São Mateus (IG)	PR	Almofada Laminada (AL)	Tradicional	MIR/NIR (PCA)
A05	São Mateus (IG)	PR	Almofada Transparente (AT)	Extra grossa	MIR/NIR (PCA)
A06	Guamiranga	PR	Vácuo Laminada (VL)	Tradicional	MIR/NIR (PCA)
A07	Abelardo Luz	SC	Almofada Laminada (AL)	Tradicional	MIR/NIR (PCA)
A08	Jaborá	SC	Vácuo Laminada (VL)	Suprema	MIR/NIR (PCA)
A09	Jaborá	SC	Papel (PP)	Tradicional	MIR/NIR (PCA)
A10	Jaborá	SC	Vácuo Laminada (VL)	Tradicional	MIR/NIR (PCA)
A11	Canoinhas	SC	Vácuo Laminada (VL)	Orgânica	MIR/NIR (PCA)
A12	Canoinhas	SC	Almofada Laminada (AL)	Nativa	MIR/NIR (PCA)
A13	Barão do Cotegipe	RS	Vácuo Laminada (VL)	Nativa	MIR/NIR (PCA)
A14	Barão do Cotegipe	RS	Vácuo Laminada (VL)	Orgânica	MIR/NIR (PCA)
A15	Barão do Cotegipe	RS	Vácuo Laminada (VL)	Premium	MIR/NIR (PCA)
A16	Barão do Cotegipe	RS	Almofada Laminada (AL)	Tradicional	MIR/NIR (PCA)

TABELA 1 - Identificação das amostras de erva-mate e principais informações fornecidas nos rótulos
(continuação)

Código	Origem	Estado	Embalagem	Tipo	Etapa
A11	Canoinhas	SC	Vácuo Laminada (VL)	Orgânica	MIR/NIR (PCA)
A12	Canoinhas	SC	Almofada Laminada (AL)	Nativa	MIR/NIR (PCA)
A13	Barão do Cotegipe	RS	Vácuo Laminada (VL)	Nativa	MIR/NIR (PCA)
A14	Barão do Cotegipe	RS	Vácuo Laminada (VL)	Orgânica	MIR/NIR (PCA)
A15	Barão do Cotegipe	RS	Vácuo Laminada (VL)	Premium	MIR/NIR (PCA)
A16	Barão do Cotegipe	RS	Almofada Laminada (AL)	Tradicional	MIR/NIR (PCA)
A17	Barão do Cotegipe	RS	Almofada Laminada (AL)	Tradicional	MIR/NIR (PCA)
A18	Barão do Cotegipe	RS	Almofada Laminada (AL)	Grossa	MIR/NIR (PCA)
A19	Machadinho	RS	Vácuo Laminada (VL)	Tradicional	MIR/NIR (PCA)
A20	Erechim	RS	Papel (PP)	Adoçada	MIR/NIR (PCA)
A21	Erechim	RS	Papel (PP)	Adoçada	MIR/NIR (PCA)
A22	Venâncio Aires	RS	Almofada Laminada (AL)	Tradicional	MIR/NIR (PCA)
A23	Venâncio Aires	RS	Vácuo Laminada (VL)	Nativa	MIR/NIR (PCA)
A24	Venâncio Aires	RS	Almofada Laminada (AL)	Premium	MIR/NIR (PCA)
A25	Venâncio Aires	RS	Almofada Laminada (AL)	Grossa	MIR/NIR (PCA)
A26	Arvorezinha	RS	Almofada Laminada (AL)	Nativa	MIR/NIR (PCA)
A27	Arvorezinha	RS	Almofada Laminada (AL)	Tradicional	MIR/NIR (PCA)
A28	Erechim	RS	Almofada Laminada (AL)	Adoçada	MIR/NIR (PCA)
A29	Santa Rosa	RS	Vácuo Laminada (VL)	Grossa	MIR/NIR (PCA)
A30	Lajeado	RS	Almofada Laminada (AL)	Reserva	MIR/NIR (PCA)
A31	Erechim	RS	Papel (PP)	Grossa	NIR (PCA e PLS-DA)
A32	Catanduvas	SC	Papel Laminado (PL)	Tradicional	NIR (PCA e PLS-DA)
A33	Jandaia do Sul	PR	Papel Laminado (PL)	Pura Folha	NIR (PCA e PLS-DA)
A34	Arvorezinha	RS	Almofada Laminada (AL)	Premium	NIR (PCA e PLS-DA)
A35	Venâncio Aires	RS	Papel Laminado (PL)	Pura folha	NIR (PCA e PLS-DA)
A36	São Mateus (IG)	PR	Almofada Laminada (AL)	Tradicional	NIR (PCA e PLS-DA)
A37	Abelardo Luz	SC	Almofada Laminada (AL)	Tradicional	NIR (PCA e PLS-DA)
A38	Campo Grande	MS	Papel (PP)	Tererê	NIR (PCA e PLS-DA)

TABELA 1 - Identificação das amostras de erva-mate e principais informações fornecidas nos rótulos
(continuação)

Código	Origem	Estado	Embalagem	Tipo	Etapa
A39	Barão do Cotegipe	RS	Vácuo Laminada (VL)	Tradicional	NIR (PCA e PLS-DA) IMAGEM TALOS
A40	Barão do Cotegipe	RS	Vácuo Laminada (VL)	Nativa	NIR (PCA e PLS-DA) IMAGEM TALOS
A41	Arvorezinha	RS	Almofada Laminada (AL)	Tradicional	NIR (PCA e PLS-DA)
A42	Erechim	RS	Papel Laminado (PL)	Adoçada	NIR (PCA e PLS-DA)
A43	Erechim	RS	Papel Laminado (PL)	Premium	NIR (PCA e PLS-DA)
A44	Arvorezinha	RS	Vácuo Laminada (VL)	Orgânica	NIR (PCA e PLS-DA)
A45	Rio das Antas	SC	Vácuo Laminada (VL)	Premium	NIR (PCA e PLS-DA)
A46	Xaxim	SC	Vácuo Laminada (VL)	Tradicional	NIR (PCA e PLS-DA) IMAGEM TALOS
A47	Sapiranga	RS	Vácuo Laminada (VL)	Nativa	NIR (PCA e PLS-DA)
A48	Arvorezinha	RS	Vácuo Laminada (VL)	Premium	NIR (PCA e PLS-DA)
A49	General Carneiro	PR	Vácuo Laminada (VL)	Nativa	NIR (PCA e PLS-DA)
A50	Catanduvas	SC	Vácuo Laminada (VL)	Premium	NIR (PCA e PLS-DA)
A51	Erechim	RS	Vácuo Laminada (VL)	Nativa	NIR (PCA e PLS-DA)
A52	Barão do Cotegipe	RS	Vácuo Laminada (VL)	Orgânica	NIR (PCA e PLS-DA) IMAGEM TALOS
A53	Canoinhas	SC	Vácuo Laminada (VL)	Orgânica	NIR (PCA e PLS-DA)
A54	São Mateus (IG)	PR	Vácuo Laminada (VL)	Nativa	NIR (PCA e PLS-DA)
A55	Machadinho	RS	Vácuo Laminada (VL)	Tradicional	NIR (PCA e PLS-DA)
A56	Ilópolis	RS	Vácuo Laminada (VL)	Tradicional	NIR (PCA e PLS-DA)
A57	Abelardo Luz	SC	Vácuo Laminada (VL)	Premium	NIR (PCA e PLS-DA)
A58	Ilópolis	RS	Vácuo Laminada (VL)	Nativa	NIR (PCA e PLS-DA)
A59	Venâncio Aires	RS	Vácuo Laminada (VL)	Requinte	NIR (PCA e PLS-DA)
A60	Encantado	RS	Papel (PP)	Tradicional	NIR (PCA e PLS-DA)
A61	Arvorezinha	RS	Vácuo Laminada (VL)	Pura Folha	NIR (PCA e PLS-DA)
A62	Abelardo Luz	SC	Almofada Laminada (AL)	Extra	NIR (PCA e PLS-DA)
A63	Jaborá	SC	Almofada Laminada (AL)	Suprema	NIR (PCA e PLS-DA)
A64	Marau	RS	Papel (PP)	Artesanal	NIR (PCA e PLS-DA)
A65	Dourados	MS	Papel (PP)	Pura Folha	NIR (PCA e PLS-DA)
A66	Lapa	PR	Papel Laminado (PL)	Tradicional	NIR (PCA e PLS-DA)

TABELA 1 - Identificação das amostras de erva-mate e principais informações fornecidas nos rótulos
(continuação)

Código	Origem	Estado	Embalagem	Tipo	Etapa
A67	Barão do Cotegipe	RS	Papel Laminado (PL)	Exportação	NIR (PCA e PLS-DA)
A68	Arvorezinha	RS	Almofada Laminada (AL)	Tradicional	NIR (PCA e PLS-DA)
A69	Encantado	RS	Papel (PP)	Exportação	NIR (PCA e PLS-DA)
A70	São Mateus do Sul	PR	Papel (PP)	Exportação	NIR (PCA e PLS-DA)
A71	Barão do Cotegipe	RS	Vácuo Laminada (VL)	Premium	NIR (PCA e PLS-DA)
A72	Arvorezinha	RS	Papel Laminado (PL)	Premium	NIR (PCA e PLS-DA) IMAGEM TALOS
A73	Encantado	RS	Almofada Laminada (AL)	Exportação	NIR (PCA e PLS-DA)
A74	Ilópolis	RS	Almofada Laminada (AL)	Adoçada	NIR (PLS-DA)
A75	Ilópolis	RS	Almofada Laminada (AL)	Adoçada	NIR (PLS-DA)
A76	Ilópolis	RS	Almofada Laminada (AL)	Adoçada	NIR (PLS-DA)
A77	Ilópolis	RS	Almofada Laminada (AL)	Adoçada	NIR (PLS-DA)
A78	Ilópolis	RS	Almofada Laminada (AL)	Adoçada	NIR (PLS-DA)
A79	Ilópolis	RS	Almofada Laminada (AL)	Adoçada	NIR (PLS-DA)
A80	Erechim	RS	Almofada Laminada (AL)	Adoçada	NIR (PLS-DA)
A81	Arvorezinha	RS	Almofada Laminada (AL)	Adoçada	NIR (PLS-DA)
B01	São Mateus (IG)	PR	Almofada Laminada (AL)	Tradicional	RMN 1D
B02	São Mateus (IG)	PR	Almofada Laminada (AL)	Tradicional	RMN 1D
B03	São Mateus do Sul	PR	Almofada Laminada (AL)	Nativa	RMN 1D
B04	São Mateus (IG)	PR	Almofada Laminada (AL)	Exportação	RMN 1D
B05	São Mateus (IG)	PR	Almofada Laminada (AL)	Tradicional	RMN 1D
B06	São Mateus do Sul	PR	Almofada Laminada (AL)	Nativa	RMN 1D e 2D
B07	São Mateus (IG)	PR	Almofada Laminada (AL)	Nativa	RMN 1D
B08	São Mateus do Sul	PR	Almofada Laminada (AL)	Tradicional	RMN 1D
B09	Santa Rosa	RS	Almofada Laminada (AL)	Tradicional	RMN 1D
B10	Santa Rosa	RS	Almofada Laminada (AL)	Tradicional	RMN 1D
B11	Arvorezinha	RS	Almofada Laminada (AL)	Nativa	RMN 1D
B12	Arvorezinha	RS	Almofada Laminada (AL)	Nativa	RMN 1D
B13	Barão do Cotegipe	RS	Almofada Laminada (AL)	Tradicional	RMN 1D

TABELA 1 - Identificação das amostras de erva-mate e principais informações fornecidas nos rótulos (conclusão)

Código	Origem	Estado	Embalagem	Tipo	Etapa
B14	Barão do Cotegipe	RS	Almofada Laminada (AL)	Tradicional	RMN 1D
B15	Portão	RS	Almofada Laminada (AL)	Premium	RMN 1D
B16	Portão	RS	Almofada Laminada (AL)	Premium	RMN 1D
B17	Canoinhas	SC	Almofada Laminada (AL)	Orgânica	RMN 1D
B18	Canoinhas	SC	Almofada Laminada (AL)	Orgânica	RMN 1D
B19	Canoinhas	SC	Almofada Laminada (AL)	Nativa	RMN 1D
B20	Canoinhas	SC	Almofada Laminada (AL)	Nativa	RMN 1D
B21	Porto Vitória	PR	Almofada Transparente (AT)	Nativa	RMN 1D
B22	Mallet	PR	Papel (PP)	Tradicional	RMN 1D
B23	São Mateus (IG)	PR	Papel Laminado (PL)	Nativa	RMN 1D
B24	São Mateus (IG)	PR	Papel Laminado (PL)	Especial	RMN 1D
B25	São Mateus (IG)	PR	Papel Laminado (PL)	Verdinha	RMN 1D
B26	São Mateus (IG)	PR	Papel Laminado (PL)	Tradicional	RMN 1D e 2D
B27	General Carneiro	PR	Papel Laminado (PL)	Nativa	RMN 1D
B28	Bituruna	PR	Papel Laminado (PL)	Descansada	RMN 1D
B29	São Mateus (IG)	PR	Papel Laminado (PL)	Tradicional	RMN 1D
B30	São Mateus do Sul	PR	Papel Laminado (PL)	Nativa	RMN 1D
B31	São Mateus do Sul	PR	Vácuo Laminada (VL)	Verdinha	RMN 1D
B32	São Mateus (IG)	PR	Vácuo Laminada (VL)	Tradicional	RMN 1D
B33	São Mateus do Sul	PR	Vácuo Laminada (VL)	Nativa	RMN 1D
B34	São Mateus (IG)	PR	Vácuo Laminada (VL)	Tradicional	RMN 1D
B35	Erechim	SC	Vácuo Laminada (VL)	Nativa	RMN 1D
B36	São Mateus (IG)	PR	Vácuo Laminada (VL)	Tradicional	RMN 1D
B37	São João do Triunfo	PR	Vácuo Laminada (VL)	Orgânica	RMN 1D
B38	Jaborá	SC	Vácuo Laminada (VL)	Suprema	RMN 1D e 2D
B39	Arvorezinha	RS	Vácuo Laminada (VL)	Tradicional	RMN 1D
B40	Arvorezinha	RS	Vácuo Laminada (VL)	Premium	RMN 1D
B41	Erechim	SC	Vácuo Laminada (VL)	Tradicional	RMN 1D
B42	Ilópolis	RS	Vácuo Laminada (VL)	Tradicional	RMN 1D
B43	Santa Lúcia	SC	Vácuo Laminada (VL)	Premium	RMN 1D
B44	Sapiranga	RS	Vácuo Laminada (VL)	Premium	RMN 1D
B45	Ilópolis	RS	Vácuo Laminada (VL)	Especial	RMN 1D
B46	Rio das Antas	SC	Vácuo Laminada (VL)	Premium	RMN 1D
B47	Arvorezinha	RS	Vácuo Laminada (VL)	Tradicional	RMN 1D
B48	Venâncio Aires	RS	Vácuo Laminada (VL)	Requinte	RMN 1D e 2D
B49	Arvorezinha	RS	Vácuo Laminada (VL)	Ouro	RMN 1D
B50	Ijuí	RS	Vácuo Laminada (VL)	Nativa	RMN 1D

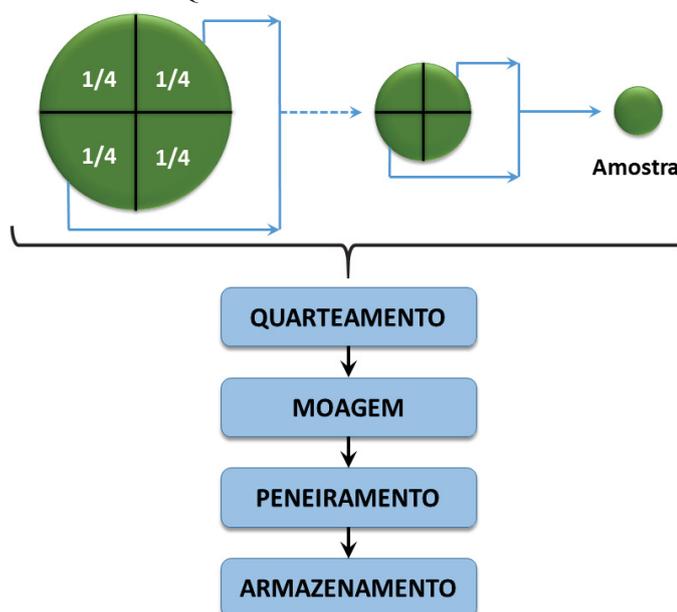
FONTE: A autora (2023).

4.2 PROCESSO DE AMOSTRAGEM

A erva-mate beneficiada pode conter até 30% de caule em sua composição e geralmente é comercializada em embalagens de 500 ou 1000 g, sendo, portanto, uma amostra bruta grande e heterogênea. Para reduzir possíveis problemas relacionados com a heterogeneidade das amostras, foi necessário efetuar o preparo das amostras brutas, a fim de obter amostras de laboratório com homogeneidade e representatividade adequadas.

Conforme apresentado na FIGURA 19, o conteúdo total das embalagens foi disposto sob uma folha de papel vegetal dentro de um recipiente circular e a amostra foi nivelada. Posteriormente, as amostras foram quarteadas de maneira sucessiva até a obtenção de uma amostra representativa de aproximadamente 250 mg. As subamostras foram trituradas, em equipamento do tipo mixer, durante 5 min, e peneiradas para separar a fração granulométrica entre 50-80 mesh. A fim de evitar degradação, as amostras foram acondicionadas em embalagens laminadas e mantidas sob refrigeração até o momento das análises.

FIGURA 19 - ETAPAS DO QUARTEAMENTO DAS AMOSTRAS DE ERVA-MATE



FONTE: Adaptado de Diniz (2013).

4.3 AQUISIÇÃO DOS ESPECTROS NA REGIÃO DO INFRAVERMELHO

As amostras de erva-mate foram submetidas à análise espectroscópica na região do infravermelho médio (MIR, do inglês *Mid-infrared*) e próximo (NIR, do inglês *Near-infrared*) em espectrômetro FTIR Vertex 70 – Bruker Optics Inc., equipado com acessório de reflectância difusa com transformada de Fourier (DRIFT) e com sonda de fibra óptica. O branco analítico

(background) foi obtido com o uso de um espelho circular. A aquisição dos espectros de reflectância (R) foi realizada à temperatura ambiente (23-25 °C) com resolução de 4 cm⁻¹ e 64 varreduras na região espectral de 10000-4000 cm⁻¹ (NIR) e de 4000-400 cm⁻¹ (MIR), sendo os resultados obtidos expressos em log (1/R). Foram registradas 3 réplicas amostrais, sendo que a cada 10 aquisições o branco analítico foi registrado novamente para garantir a qualidade dos espectros com relação às variações na linha de base. A obtenção dos espectros com a sonda foi realizada a fim de verificar a potencialidade de simplificação da coleta de dados.

Os espectros foram exportados na extensão .dpt e importados para ambiente MATLAB, versão 2018a (MathWorks, Natick, EUA) onde foram processados e analisados com o pacote PLS Toolbox, versão 8.7. Os dados foram organizados em matrizes (X_{NIR} e X_{MIR}) que em seguida foram utilizadas nas análises quimiométricas.

4.4 AQUISIÇÃO DOS ESPECTROS DE RMN HR-MAS DE ¹H

Os espectros de RMN HR-MAS de ¹H foram adquiridos em um espectrômetro Bruker Avance II 400 (9,4 T) observando o núcleo de ¹H a 400 MHz, equipado com uma sonda quadrinuclear (¹H; ¹³C; ¹⁵N, ²H) de alta resolução de 4 mm MAS, com gradiente de campo na direção do ângulo mágico ($\theta = 54,74^\circ$) a 298 K. A sequência de pulsos zgpr foi utilizada para a pré-saturação do sinal residual de água. O ângulo mágico (MAS) foi ajustado diariamente, enquanto que a sintonia (*matching* e *tunning*) e o *shimming* foram ajustados para cada amostra.

Os espectros foram registrados com 256 transientes, tempo de aquisição de 4,0 s, com 64 K pontos durante a aquisição, utilizando janela espectral de 8012,820 Hz, e ganho do receptor de 114. Da mesma forma, os espectros foram referenciados em relação ao sinal do TMS em δH 0,00 e a velocidade de giro do rotor setada em 5 kHz. Além disso, a calibração do pulso de 90° foi otimizada para $pw90 = \frac{1}{4} pw360$ e $p1 = 5,15$ Ps. Para a obtenção dos espectros, 10 mg de amostra e 45 μ L do solvente CD₃OD (0,05% de TMS) foram acondicionados em um rotor de ZrO₂ semiesférico, com um volume de 50 μ L. A pesagem das amostras foi realizada em uma balança semi-microanalítica, modelo AUW 220D (Shimadzu, Kyoto, Japão), com precisão de 0,01 mg e capacidade máxima de 220 g.

Os espectros foram pré-processados com o software TopSpin, versão 3.2 (Bruker BioSpin Inc.). As fases foram corrigidas manualmente e as linhas de base automaticamente. As análises foram realizadas em duplicata para um conjunto de 50 amostras (TABELA 1, amostras B01 até B50), resultando em um conjunto amostral organizado na matriz X_{RMN} que foi posteriormente importada em ambiente MATLAB para a etapa de análises quimiométricas.

4.5 AQUISIÇÃO DAS IMAGENS DIGITAIS

As imagens digitais foram adquiridas em um scanner de mesa da marca CANON, modelo LiDE 220, com resolução de 4800 dpi. O dispositivo utilizado para a captura das imagens digitais possui tampa superior de dois níveis (FIGURA 20a), que permite acomodar itens de maior espessura conforme mostrado na FIGURA 20b. As amostras foram acondicionadas em placas Petri de cultura de células (35x10mm) possibilitando a análise de 24 poços por imagem. Para reduzir interferências provenientes de FONTES difusas de iluminação, o scanner foi adaptado com a inclusão de um anteparo para fechamento completo das laterais. Além disso, um molde (FIGURA 20c) foi utilizado para assegurar o distanciamento entre as placas, visto que quando colocadas muito próximas as placas de vidro causam reflexos entre si.

FIGURA 20 - APARATO INSTRUMENTAL UTILIZADO NA OBTENÇÃO DAS IMAGENS DIGITAIS



FONTE: A autora (2023).

4.5.1 Processamento das imagens digitais

As imagens capturadas foram armazenadas em formato .png (*portal network graphic*) para possibilitar a separação matemática dos canais de interesse. O processamento das imagens foi realizado em ambiente MATLAB, utilizando as ferramentas PLS_Toolbox, Image Processing Toolbox e rotinas desenvolvidas no laboratório DATA Lab que possibilitaram a automatização das etapas de processamento para posterior análise quimiométrica: normalização e equalização de histogramas, identificação da posição das placas, a segmentação e individualização das placas e a posterior seleção da região de interesse (ROI, do inglês *Region of interest*) para obtenção dos histogramas de cor (GONZALES, 2018).

4.5.1.1 Correção da luminosidade

Inicialmente, a imagem foi convertida da escala RGB para a escala de tons de cinza, através da função `rgb2gray`, disponível no Imaging Process toolbox. Em seguida, foram aplicados três ajustes para realce de contraste: a) a **normalização** foi obtida pela função `imadjust`, a qual normaliza os valores dos pixels da imagem por um valor de intensidade entre zero e um; b) a **equalização** de histograma foi obtida pela função `histeq`, que ajusta os valores de pixels da imagem de acordo com um histograma de 64 divisões e distribuição contínua de valores; c) e a **equalização adaptativa** que foi obtida pela função `adapthisteq`, que também realiza o realce da imagem a partir de um histograma, porém seleciona pequenas partes da imagem por iteração para realizar um histograma de forma local.

4.5.1.2 Segmentação e individualização

Após a etapa de correção de luminosidade, a rotina para reconhecimento automático dos poços, desenvolvida no DATA Lab (FIGURA 21), foi aplicada aos dados de imagem. A rotina utiliza uma expressão lógica baseada na transformada circular de Hough, que detecta as bordas de um círculo a partir de seu raio e do contraste entre o círculo e o fundo da imagem (GONZALES, 2018). A rotina, utiliza a transformada circular de Hough já implementada no Image Processing Toolbox pela função `imfindcircles` e testa diferentes níveis de sensibilidade do algoritmo até que o número de círculos identificados seja igual ao número de poços presentes na imagem analisada (FIGURA 22). O mapeamento da localização das ROI em cada amostra (círculos em vermelho) é então utilizado como máscara na imagem original (FIGURA 23), o que permite que apenas os pixels presentes dentro da ROI sejam considerados para a amostra.

FIGURA 21 - ROTINA DE RECONHECIMENTO DOS POÇOS

```
centers = [];
Sens = 0;

for i = 0.95:0.001:0.99
    if size(centers,1)~=nplts
        Sens = 1;
        [centers,radii] = imfindcircles(im_imadjust,rad,'ObjectPolarity','dark','Sensitivity',Sens);
    elseif size(centers,1) > nplts
        break;
    end
    disp(Sens)
end
```

FONTE: DATA Lab (2023).

FIGURA 22 - ROTINA DE APLICAÇÃO DA MÁSCARA DE IDENTIFICAÇÃO DOS POÇOS

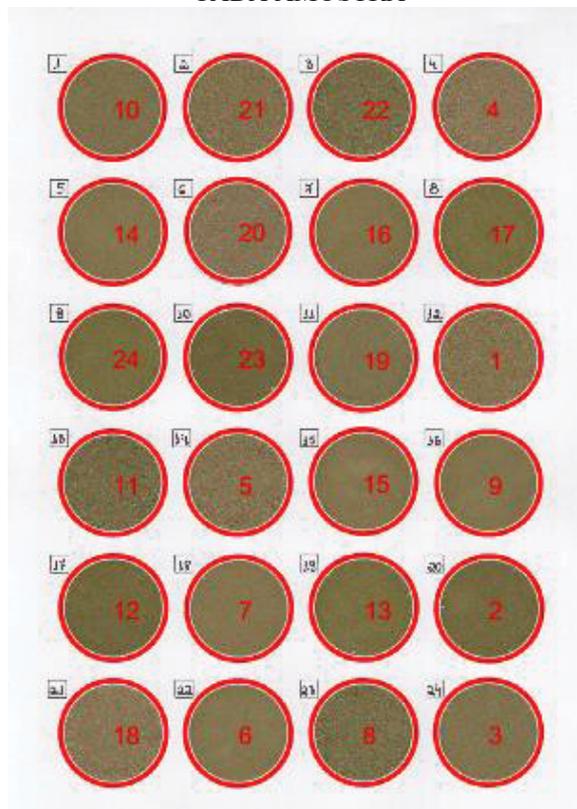
```

figure;
imshow(im_sel)
well = 0;
h = viscircles(centers, radii); % Mostrar círculos na imagem
well_sup = zeros(nplts,1);
% Gerar texto com número de identificação das placas
for l=1:nplts
    text(centers(1,1),centers(1,2),num2str(1+well),'Color','r')
    well_sup(1) = 1+well;
end
centerord = sortrows(sortrows([well_sup fix(centers./100) centers],3),2);
centerord(:,2:3) = [];

```

FONTE: DATALab (2023).

FIGURA 23 - IDENTIFICAÇÃO DA POSIÇÃO DOS POÇOS E DAS REGIÕES DE INTERESSE PARA CADA AMOSTRA



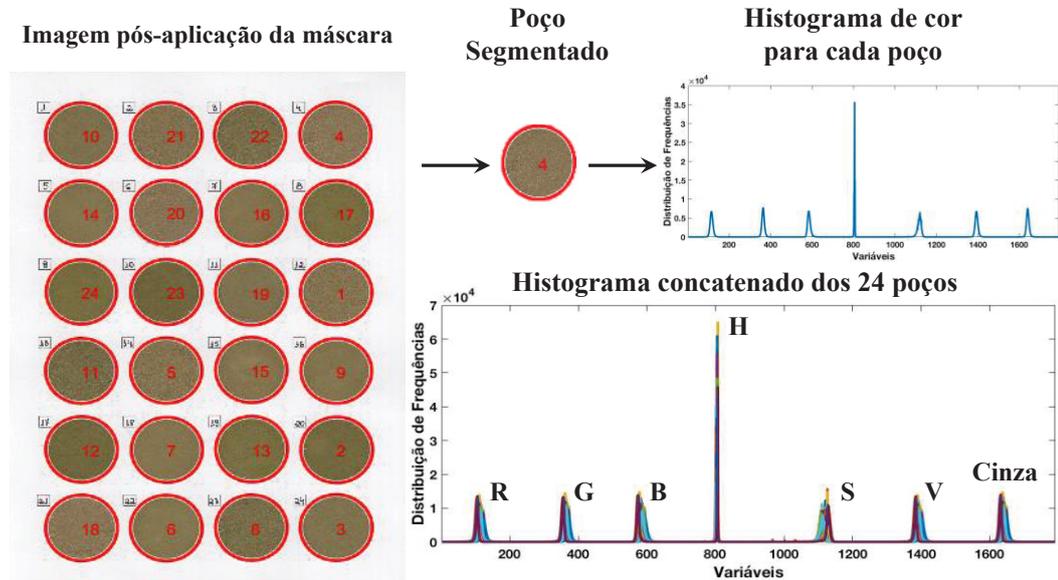
FONTE: A autora (2023).

4.4.1.3 Obtenção dos histogramas de cor

Após a aplicação da máscara, a imagem em RGB foi convertida para os modelos de cor HSV e HSI e a separação dos canais de interesse (RGB, HSV e tons de cinza) foi efetuada. Devido às configurações do aparato de obtenção de imagens, a partir de cada imagem foi possível obter os histogramas de cor de 24 poços. Conforme discutido na seção 2.7, os histogramas de cada canal de cor são compostos por 256 divisões (variações de tom de 0 a 255).

Sendo assim, as informações concatenadas de todos os canais (RGB+HSV+cinza) resultou em um histograma com 1792 variáveis para cada amostra (FIGURA 24).

FIGURA 24 - PROCESSO DE OBTENÇÃO DOS HISTOGRAMAS DE COR A PARTIR DA ROTINA DE RECONHECIMENTO DE POÇOS



FONTE: A autora (2023).

3.7.3 Preparo das misturas com adição de talos

Para o desenvolvimento da metodologia para quantificação de talos (obtidos via peneiração de amostras comerciais) com base em imagens digitais, foram preparadas misturas para simular a adição de 5-60% (com acréscimos de 5%) de talos em amostras de erva-mate, sendo que a faixa foi selecionada com base na legislação não-vigente de classificação de amostras baseada na quantidade de talo presente. As misturas foram acondicionadas nas placas de Petri para obtenção das imagens digitais no aparato descrito na seção 3.7. A fim de obter um modelo robusto, 5 amostras representativas do conjunto (indicadas na TABELA 1) foram selecionadas por meio do algoritmo *Kennard-Stone* para serem utilizadas na preparação das misturas em triplicata. Da mesma forma, o talo utilizado foi uma amostra composta por talos retirados das amostras comerciais. Assim como as amostras, os talos passaram por uma etapa de homogeneização prévia em mixer. Os dados foram organizados em uma matriz (X_{TAL}) que em seguida foi utilizada nas análises quimiométricas.

5 RESULTADOS E DISCUSSÃO

5.1 QUIMIOMETRIA APLICADA À DADOS DE ESPECTROSCOPIA MIR E NIR

5.1.1 Análise exploratória comparativa – regiões MIR e NIR

Tendo conhecimento de que, as absorções observadas na região do NIR são menos intensas e resultam em bandas de absorção menos pronunciadas quando comparadas com os espectros na região MIR, esse estudo foi iniciado com uma análise exploratória comparativa entre as duas faixas espectrais. O intuito dessa etapa foi avaliar qual das regiões apresentaria maior potencial para aplicação em metodologias com foco na discriminação das amostras de erva-mate de acordo com a origem geográfica.

5.1.1.1 Infravermelho médio (MIR)

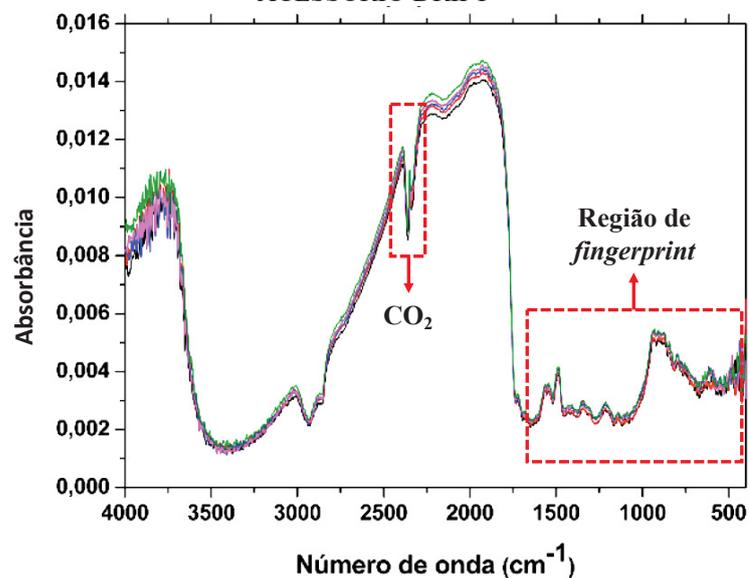
O perfil espectral na região MIR das amostras de erva-mate obtido com acessório DRIFT na faixa de 4000-450 cm^{-1} , é apresentado na FIGURA 25. A banda forte e larga na região de 3400 cm^{-1} está associada a ligações do tipo O-H provenientes de álcoois, água e fenóis (CHRISTOU et al., 2018). As bandas que aparecem entre 3000 cm^{-1} e 2800 cm^{-1} são atribuídas a estiramentos de ligações C-H. A banda em torno de 2900 cm^{-1} é atribuída ao estiramento simétrico de CH_3 associada principalmente a presença de lipídios e flavonoides, enquanto que a banda em torno 2850 cm^{-1} é atribuída ao estiramento simétrico de CH_2 associada principalmente a presença de lipídios e cafeína. A banda na região de 2350 cm^{-1} é associada a deformação axial assimétrica da molécula de CO_2 (SILVERSTEIN, 2006). De acordo com a literatura, a banda na região de 1630 cm^{-1} pode ser atribuída a deformações de ligações N-H e estiramento de ligações C-N em compostos do grupo amida.

Na região de impressão digital (*fingerprint*) observam-se absorções mais fracas, porém bastante úteis para a identificação de compostos em matrizes vegetais (DE SOUZA e POPPI, 2012). De acordo com estudos de amostras similares, nessa região são observadas bandas relacionadas a estiramentos e deformações de ligações C=O, O-H, C-O, C-O-C, COO^- , C-N, CH_2 e CH_3 associadas a flavonóides, clorofilas, lipídios, proteínas, álcoois, fenóis, ésteres e amidas (CHRISTOU et al., 2018; SAKANE et al., 2019).

Uma vez que os espectros da região MIR apresentam efeito mais acentuado da banda de CO_2 , uma seleção de variáveis a priori foi efetuada pela eliminação (corte) dessa região espectral. Mesmo após essa etapa, regiões com elevado ruído instrumental ou sem informação espectral continuavam sendo incorretamente consideradas significativas na avaliação

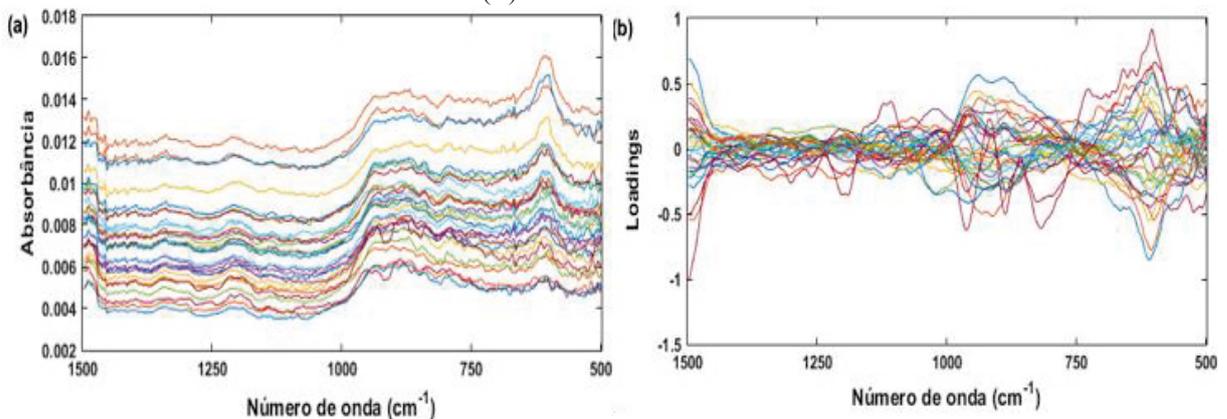
quimiométrica e por esse motivo optou-se por trabalhar apenas na região de *fingerprint* (FIGURA 26a). Os espectros de reflectância difusa sofrem efeitos de espalhamento de luz que podem ser resultado de diversos parâmetros, tais como, variações no tamanho das partículas e no empacotamento da amostra (FERRÃO, 2001). Com o intuito de reduzir o efeito das variações indesejáveis, vários pré-processamentos, (Suavização (SG) e Derivada Savitzky–Golay (DSG), Correção Multiplicativa de Espalhamento (MSC), padronização Normal de Sinal (SNV), Dados Centrados na Média (MC)) e combinações destes, foram avaliados.

FIGURA 25 - PERFIL ESPECTRAL DE AMOSTRAS DE ERVA-MATE NA REGIÃO MIR OBTIDOS COM ACESSÓRIO DRIFT



FONTE: A autora (2023).

FIGURA 26 - REGIÃO DE *FINGERPRINT* MIR DAS 30 AMOSTRAS DE ERVA-MATE ANTES (A) E APÓS (B) PRÉ-PROCESSAMENTO



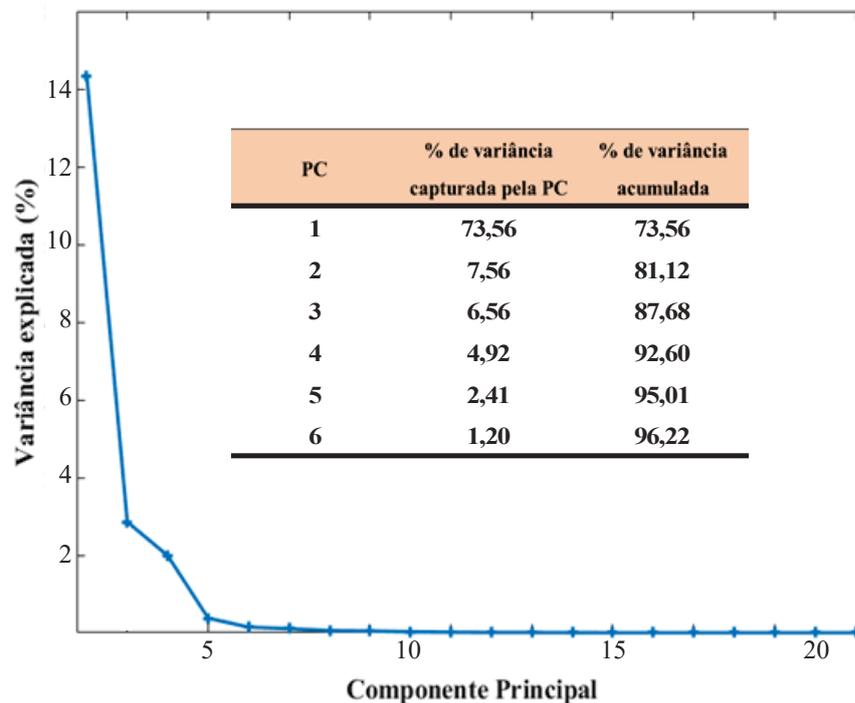
FONTE: A autora (2023).

A PCA foi efetuada em caráter exploratório visando avaliar a existência de similaridades e agrupamentos das amostras. Inicialmente, os dados foram centrados na média para que as diferenças entre os espectros fossem evidenciadas, entretanto os resultados obtidos não foram

satisfatórios, pois os espectros pré-processados apresentavam regiões bastante ruidosas e deslocamentos de linha de base (FIGURA 26a) devido aos efeitos de espalhamento. Após a realização de diversos testes com diferentes processamentos espectrais, a melhor discriminação das amostras de erva-mate foi observada com a combinação de pré-processamentos (SNV+MC+SG) que resultou no perfil espectral mostrado na FIGURA 26b. Como pode ser observado, problemas com o espalhamento dos espectros foram solucionados e uma diminuição dos ruídos foi observada. Além disso, os espectros pré-processados apresentam bandas mais bem definidas o que facilita a visualização das regiões que contém informações relevantes para a discriminação das amostras.

A relação entre as PC's e seus autovalores (FIGURA 27) mostra que as 6 primeiras PC's acumularam mais de 96% da variabilidade total existente nos dados. Como a adição de mais PC's não resultou em agrupamentos mais satisfatório das amostras em relação a origem geográfica, apenas as 6 foram mantidas.

FIGURA 27 - AUTOVALORES EM FUNÇÃO DO NÚMERO DE PC'S PARA DADOS MIR.
INSERIDO: tabela apresentando as % de variância capturada por cada PC

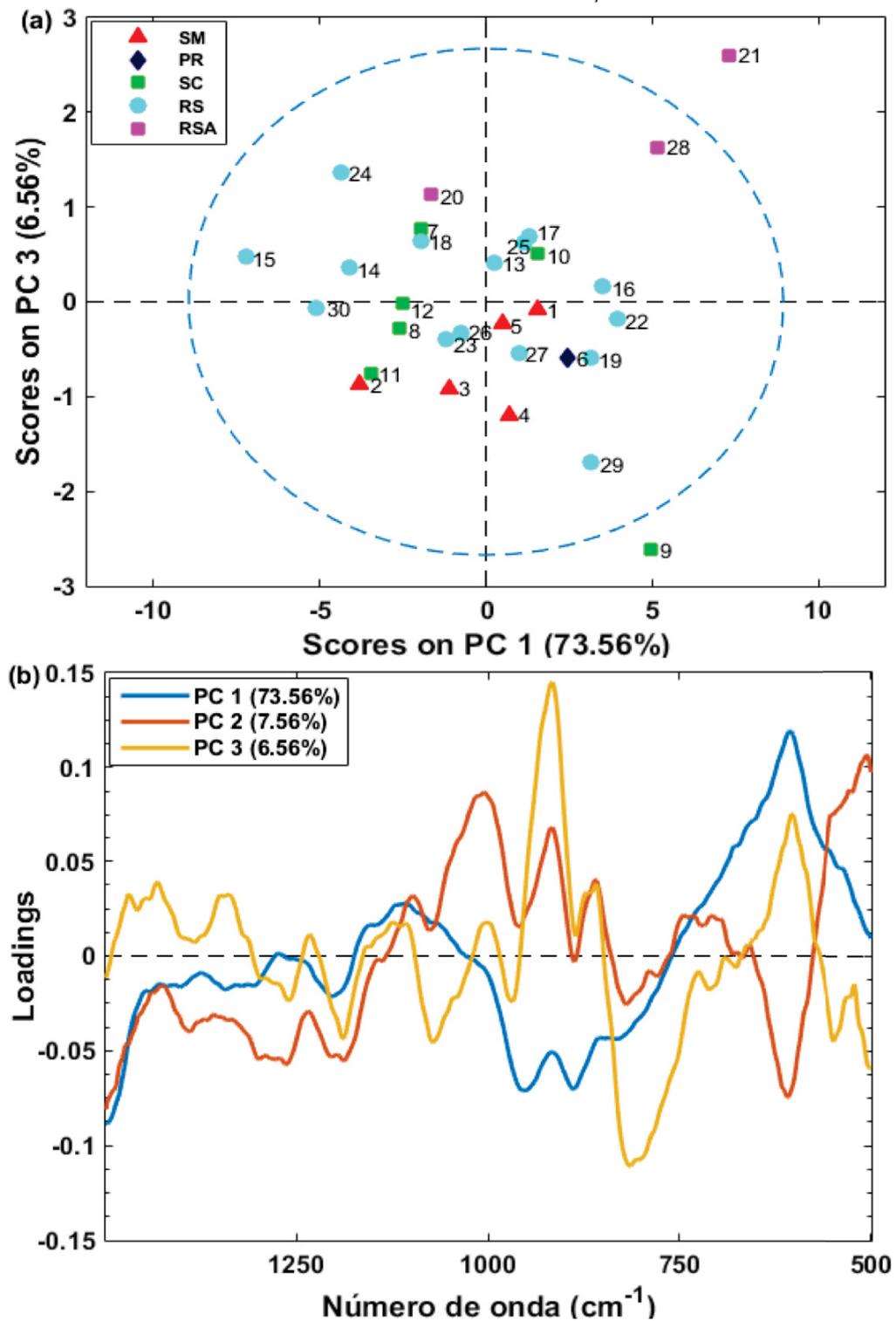


FONTE: A autora (2023).

Após a seleção do número de PC's, todas as possíveis combinações das mesmas foram avaliadas em busca de tendências de separação e agrupamento das amostras de erva-mate de acordo com as características das mesmas. Entretanto, a combinação de PC1 e PC3 foi a que apresentou as melhores tendências de agrupamento (FIGURA 28). Visando a melhor

visualização nos gráficos de *scores*, as amostras foram identificadas com cores distintas de acordo com a região de produção ou com a presença de açúcar: São Mateus do Sul (vermelho), Paraná (azul escuro), Santa Catarina (verde), Rio Grande do Sul (azul claro) e adoçadas (rosa).

FIGURA 28 - GRÁFICOS DE *SCORES* (PC1 VS PC3) (A) E *LOADINGS* (B) OBTIDO A PARTIR DOS ESPECTROS MIR DAS AMOSTRAS COMERCIAIS DE ERVA-MATE IDENTIFICADAS DE ACORDO COM A ORIGEM GEOGRÁFICA E A PRESENÇA DE SACAROSE



FONTE: A autora (2023).

Na análise do gráfico de *scores* da PC1 versus PC3 (FIGURA 28a) foram observadas tendências de separação das amostras de erva-mate de acordo com a origem geográfica. Analisando a PC3 é possível verificar que todas as amostras produzidas em São Mateus estão localizadas no lado negativo e próximas entre si. A amostra A06 (Paraná) também está deslocada para valores negativos de PC3, resultado que pode ser atribuído ao fato destas amostras serem produzidas em locais com características de solo e clima similares.

No gráfico de *loadings* (FIGURA 28b) é possível observar que as bandas que apresentam maior influência na discriminação das amostras por PC3 se encontram na faixa entre 1200-1100 cm^{-1} e em torno de 800 cm^{-1} que podem ser atribuídas a estiramentos simétricos e assimétricos de ligações C-O, provenientes de éter cíclicos associados a compostos fenólicos (ABBAS et al., 2017; LI et al., 2015). Estudos indicam que a erva-mate produzida em áreas sombreadas tende a apresentar maiores teores de compostos fenólicos e flavonóides e que a qualidade do produto é positivamente correlacionada com a concentração destes compostos (BRAGHINI et al., 2014; CRAVO FILHO et al., 2017; DA SILVEIRA et al., 2016).

No gráfico de *scores* também pode ser observado que as amostras adoçadas se encontram no lado positivo de PC3. Observando o gráfico de *loadings* vemos que as amostras localizadas nesse lado de PC3 recebem influência de bandas de absorção na região de aproximadamente 850 cm^{-1} que são associadas a vibrações C-C e C-O. De acordo com a literatura, essas bandas de absorção podem ser atribuídas a moléculas de carboidratos, incluindo a sacarose (CADET, 1999; DUARTE et al., 2002). A posição das amostras adoçadas no gráfico de *scores* indica que as mesmas também sofrem influência significativa de PC1, sendo que as amostras 21 e 28 estão deslocadas para o valores positivos, enquanto que a amostra 20 se encontra no lado negativo da mesma PC. As bandas de absorção na região entre 750-500 cm^{-1} associadas a presença de anéis aromáticos são as que mais influenciam a separação no lado positivo de PC1, além disso a região entre 1200-1100 cm^{-1} associada a presença de açúcares também apresenta influência com menor intensidade.

A combinação de PC1 e PC3 também possibilitou a visualização da tendência de agrupamento entre as amostras que são comercializadas como Premium ou Reserva Premium (A15, A24 e A30). As três amostras apresentam valores negativos em PC1 e positivos na PC3. As bandas de absorção que mais influenciam o lado negativo de PC1 são localizadas entre 1000-875 cm^{-1} , que podem ser atribuídas a ligações associadas a presença de sacarose (DUARTE et al., 2002; CADET, 1999), e entre 1500-1200 cm^{-1} que são atribuídas a deformações assimétricas de CH_3 e CH_2 . A erva-mate do tipo Premium passa por um processo de secagem em temperaturas mais baixas e diferenciado, no qual o produto não entra em contato com a

fumaça, que causa alterações nas propriedades organolépticas da erva-mate. Vale ressaltar que a amostra A14 (proveniente do mesmo local que a A15 e comercializada como orgânica) apresenta características similares as amostras do tipo Premium.

5.1.1.2 Infravermelho próximo (NIR)

A PCA aplicada aos dados NIR das 30 primeiras amostras gerou tendências de distribuição dos agrupamentos equivalentes às obtidas com os dados MIR. Esse resultado indica que, apesar de apresentar bandas menos definidas e de menor intensidade, a espectroscopia NIR pode ser utilizada para a finalidade desse estudo sem que haja prejuízo na interpretação dos resultados das análises exploratórias.

Sendo assim, devido às vantagens apresentadas pela técnica NIR, principalmente com relação à menor suscetibilidade à influência da umidade, a maior capacidade de penetração em amostras sólidas e o crescimento no interesse de aplicação da técnica na indústria de alimentos, optou-se pela continuidade das avaliações utilizando apenas os espectros na região do infravermelho próximo obtido com acessório DRIFT. Sendo assim, 43 novas amostras de erva-mate foram adquiridas para incrementar o banco de dados.

O aumento no número de amostras avaliadas é bastante relevante, principalmente quando se tratam de amostras com alta variabilidade natural, como a erva-mate. Um maior conjunto amostral aumenta a representatividade da população, conseqüentemente aumentam a capacidade de alcançar conclusões mais acuradas dentro da região de estudo e a capacidade de generalização dos resultados. Além disso, um maior número de amostras auxilia na confirmação da presença de *outliers* nos dados (FABER & FONSECA, 2022).

5.1.2 Análise exploratória via PCA – região NIR

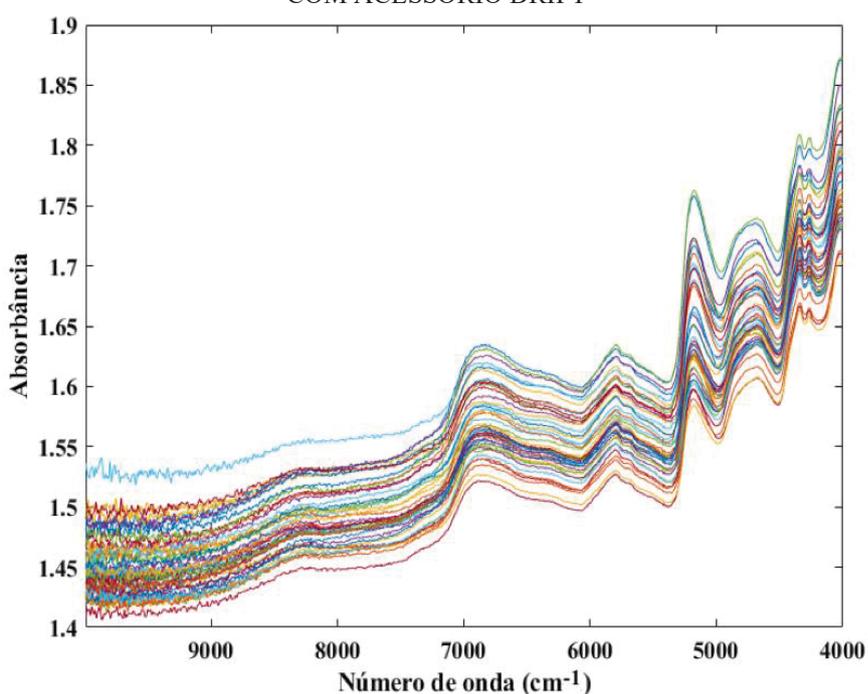
O perfil espectral na região NIR das amostras de erva obtido com o acessório DRIFT, é apresentado na FIGURA 29. As ocorrências espectrais nessa região são caracterizadas por sobreposições e combinações de absorções que podem ser atribuídas a ligações envolvendo átomos de hidrogênio, que por apresentarem pequena massa molar e grande momento de dipolo conseguem vibrar com maior amplitude (COSTA FILHO e POPPI, 2002).

A região entre 10000 e 7500 cm^{-1} é atribuída ao terceiro sobretom de ligações C-H. As bandas presentes na região entre 7500-5000 cm^{-1} indicam a presença de água nas amostras e são associadas ao primeiro sobretom de compostos com ligações C-H, C-O, O-H e C-H. Bandas

abaixo de 4500 cm^{-1} são atribuídas a bandas de combinação envolvendo ligações N-H, C-H e C-C (WORKMAN JR. & WEYER, 2012).

Uma seleção de variáveis *a priori* foi efetuada para eliminar a região espectral entre $10000\text{-}7500\text{ cm}^{-1}$, devido ao fato de não serem observadas bandas e pela presença de ruído instrumental que afeta a interpretação das variáveis significativas. Além disso, diferentes processamentos espectrais e combinações desses foram avaliados com o intuito de reduzir os efeitos de espalhamento de luz e deslocamentos de linha de base nos espectros.

FIGURA 29 - PERFIL ESPECTRAL DAS AMOSTRAS DE ERVA-MATE NA REGIÃO NIR OBTIDOS COM ACESSÓRIO DRIFT

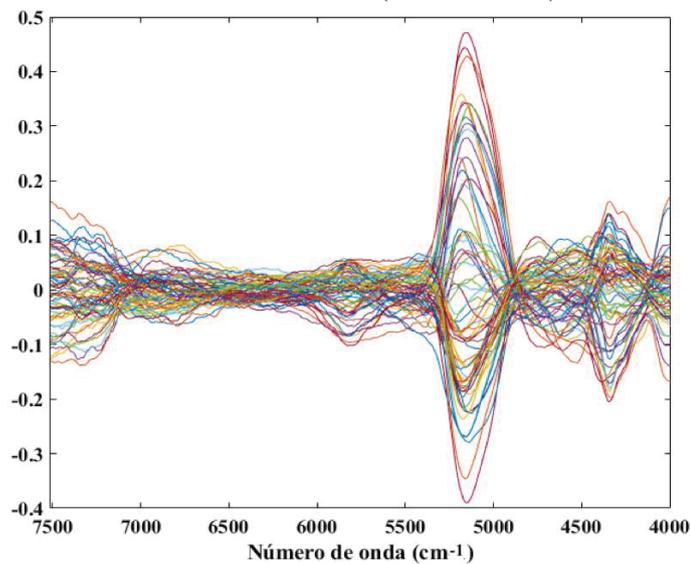


FONTE: A autora (2022).

Após a realização de testes com diferentes processamentos espectrais, foi observado que a combinação de processamentos Padronização Normal de Sinal (SNV, do inglês *Standard Normal Variate*), Dados Centrados na Média (MC, do inglês *Mean Center*) e Suavização Savitzky-Golay (SG, do inglês *Savitzky-Golay Smoothing*), que resultou no perfil espectral da FIGURA 30, solucionou problemas com o espalhamento dos espectros e minimizou ruídos. Além disso, os espectros processados apresentam bandas melhor definidas que facilitam a visualização das regiões que contém informações relevantes para a discriminação das amostras. É importante salientar, que não foi possível verificar agrupamentos de acordo com a origem geográfica das mesmas. Por esse motivo, outras possibilidades de agrupamento foram avaliadas, como: tipo de amostra, categoria de qualidade do produto e tipo de embalagem.

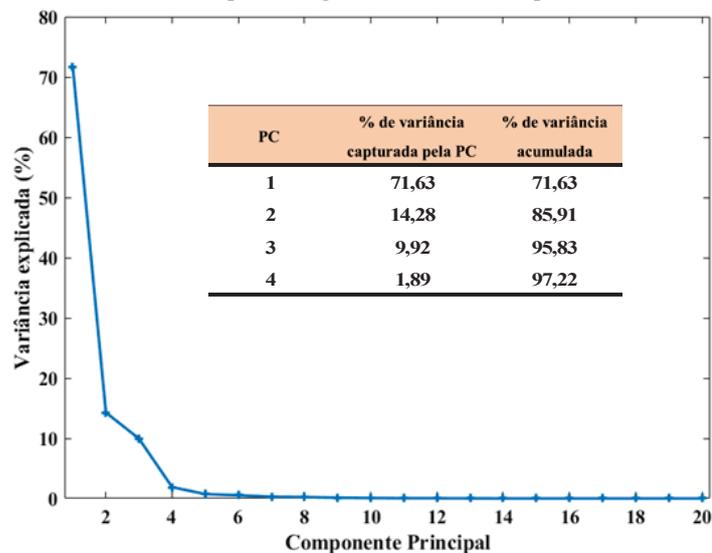
A relação entre as componentes principais (PC's) e a porcentagem de variância explicada (FIGURA 31) mostra que as 4 primeiras PC's acumularam mais de 97% da variabilidade total existente nos dados. Como a adição de um maior número de PC's não resultou em um melhor agrupamento das amostras em relação aos parâmetros avaliados, apenas essas 4 foram avaliadas.

FIGURA 30 - PERFIL ESPECTRAL NA REGIÃO NIR DAS AMOSTRAS DE ERVA-MATE APÓS PRÉ-PROCESSAMENTO (SNV+MC+SG)



FONTE: O autor (2022).

FIGURA 31 - VARIÂNCIA EXPLICADA EM FUNÇÃO DO NÚMERO DE PC'S PARA DADOS NIR.
INSERIDO: tabela de porcentagem de variância capturada em cada PC



FONTE: O autor (2022).

Visando a melhor visualização dos agrupamentos no gráfico de *scores*, as amostras foram identificadas com cores distintas: não adoçadas (azul) e adoçadas (rosa). No gráfico de

scores de PC1 *versus* PC2 (FIGURA 32a) pode ser observado que as amostras adoçadas se encontram deslocadas para o lado positivo de PC2. No gráfico de *loadings* para PC2 (FIGURA 32b) é possível constatar que as amostras localizadas nessa posição sofrem influência de bandas de absorção em duas regiões distintas: entre 7000-6000 cm^{-1} e entre 5000-4600 cm^{-1} .

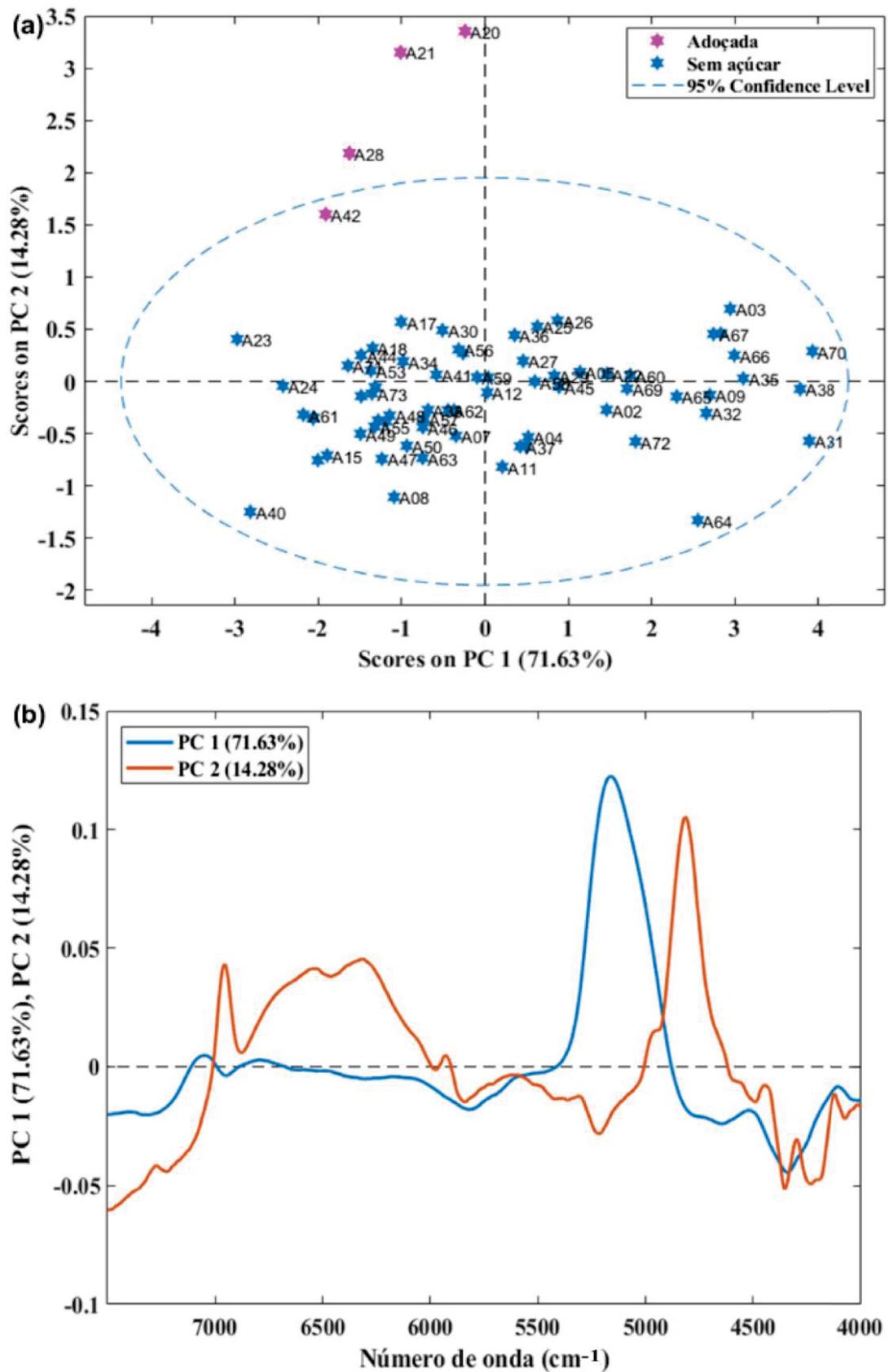
De acordo com a literatura, bandas entre 7000-6000 cm^{-1} são atribuídas a compostos contendo múltiplas ligações O-H, como é o caso da sacarose. A região entre 5000-4600 cm^{-1} é atribuída a bandas de combinação de ligações O-H e C-O, provenientes de açúcares e amidos (WORKMAN JR. & WEYER, 2012). Esse resultado corrobora a hipótese de que a técnica de infravermelho próximo pode ser utilizada para o desenvolvimento de metodologias de discriminação entre amostras de erva-mate adoçadas e sem açúcar.

As amostras adoçadas apresentaram altos valores de *leverage*, indicando que exercem influência significativa na PCA. Essas amostras foram removidas do conjunto de dados para que o efeito das outras variáveis pudesse ser avaliado. No gráfico de *scores* de PC1 *versus* PC2 (FIGURA 33) pode ser observado que, após a remoção das amostras adoçadas, a amostra A64 aparece deslocada para valores positivos de PC2 pela influência significativa da banda de água. O deslocamento da amostra pode estar atrelado a diferenças na técnica de produção, visto que é produzida artesanalmente o que pode ter contribuído para um perfil químico diferenciado. Além disso, a amostra era originalmente armazenada em uma embalagem de papel, o que pode ter ocasionado o aumento da umidade na amostra devido à alta permeabilidade.

No gráfico de *loadings* para PC2 (FIGURA 34) é possível observar que as amostras localizadas no lado positivo de PC2 sofrem influência de bandas de absorção entre 5300-5000 cm^{-1} e entre 4400-4000 cm^{-1} . Os sinais entre 5300-5000 cm^{-1} podem ser relacionados com a presença de água e de grupamentos amida, compostos fosforados, polifenóis e ésteres (WORKMAN JR. & WEYER, 2012). Enquanto entre 4400-4000 cm^{-1} podem ser associados à grupamentos contendo ligações CH, CH₂ e CH₃ presentes em diversos compostos presentes na erva-mate, como saponinas, xantinas e teobrominas (MARCELO, 2014).

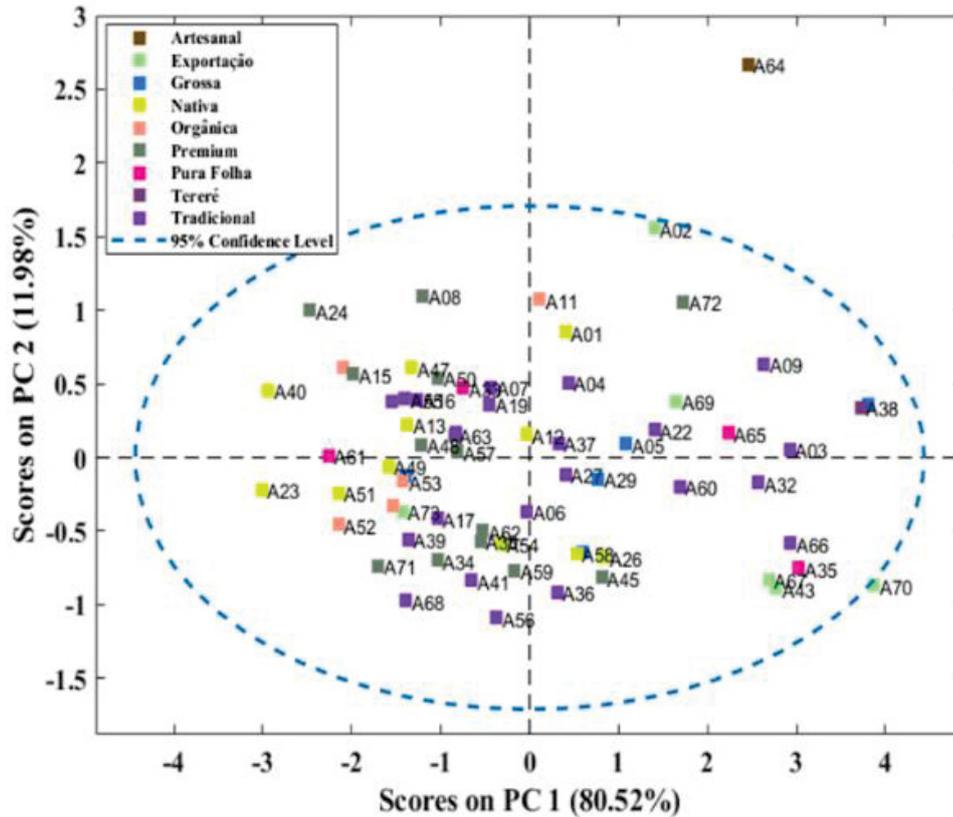
Apesar de não ocorrer um claro agrupamento das amostras de acordo com a categoria, verifica-se no gráfico de *scores* que existe uma tendência de agrupamento das amostras Nativas, Premium e Orgânicas para valores negativos de PC1 (influenciadas pelas bandas na região entre 4400-4000 cm^{-1}) e das amostras tipo exportação para valores positivos de PC1 (influenciadas pelas bandas na região entre 5300-5000 cm^{-1}).

FIGURA 32 - GRÁFICOS DE SCORES (PC1 VS PC2) (A) E LOADINGS (B) OBTIDO A PARTIR DOS ESPECTROS NIR DAS AMOSTRAS COMERCIAIS DE ERVA-MATE IDENTIFICADAS DE ACORDO COM A PRESENÇA DE SACAROSE



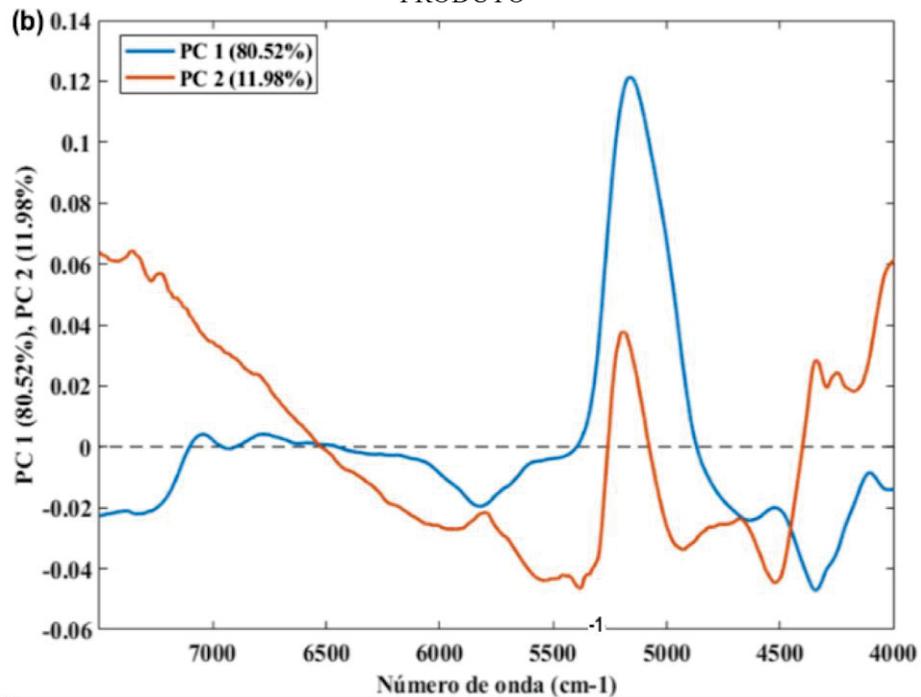
FONTE: A autora (2022).

FIGURA 33 – GRÁFICO DE SCORES (PC1 VS PC2) OBTIDO A PARTIR DOS ESPECTROS NIR DAS AMOSTRAS COMERCIAIS DE ERVA-MATE IDENTIFICADAS DE ACORDO COM O TIPO DE PRODUTO



FONTE: A autora (2023).

FIGURA 34 - GRÁFICO DE LOADINGS (PC1 VS PC2) OBTIDO A PARTIR DOS ESPECTROS NIR DAS AMOSTRAS COMERCIAIS DE ERVA-MATE IDENTIFICADAS DE ACORDO COM O TIPO DE PRODUTO

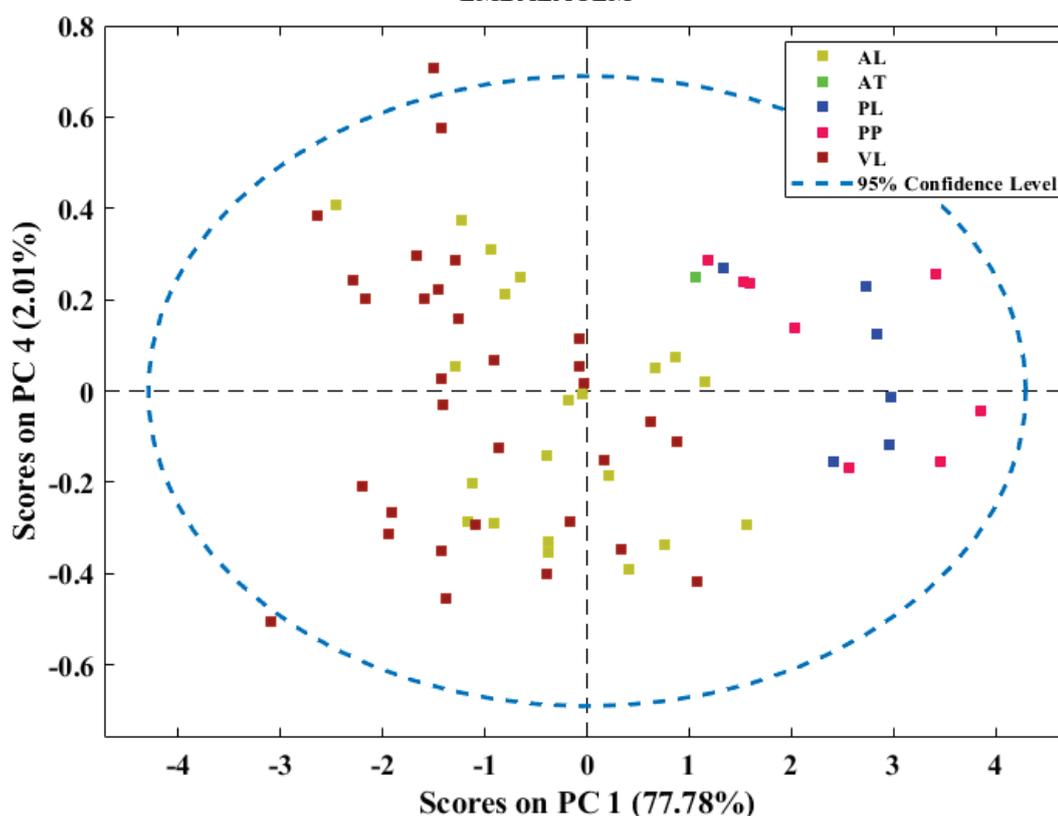


FONTE: A autora (2022).

Na combinação entre PC1 e PC4 (FIGURA 35) foi possível avaliar a distribuição das amostras de erva-mate de acordo com o tipo de embalagem. Para melhorar a visualização, o esquema de cores das mesmas foi alterado para corresponder diretamente aos diferentes tipos de embalagem utilizados nas amostras comerciais de erva-mate, sendo: almofadas laminadas (AL), almofadas transparentes (AT), papel laminado (PL), papel (PP) ou vácuo laminado (VL).

No gráfico de *scores* de PC1 versus PC4, foi possível observar a formação de dois agrupamentos: um contendo amostras que eram originalmente armazenadas em embalagens laminadas outro contendo as amostras que possuíam embalagens derivadas de papel. O aumento da umidade nas amostras embaladas em papel se dá devido à capacidade reduzida desse material de barrar a permeação de vapores de água, e está alinhado com o fato de que a banda associada à presença de água nas amostras (5300-5000 cm^{-1}) fornece a contribuição mais significativa para a discriminação das amostras para valores positivos de PC1 (FIGURA 34). Vale salientar que a amostra que possuía embalagem plástica transparente também se encontra deslocada para valores positivos de PC1, indicando que esse tipo de embalagem também apresenta menor resistência ao aumento da umidade nas amostras.

FIGURA 35 - GRÁFICOS DE *SCORES* (PC1 VS PC4) OBTIDO A PARTIR DOS ESPECTROS NIR DAS AMOSTRAS COMERCIAIS DE ERVA-MATE IDENTIFICADAS DE ACORDO COM O TIPO DE EMBALAGEM



FONTE: A autora (2022).

Apesar da associação entre a espectroscopia NIR e métodos quimiométricos ter se mostrado uma ferramenta com resultados promissores para a determinação de compostos bioativos (ROSSI & LOZANO, 2020; LIMA et al., 2019), a identificação da presença de sacarose (VIEIRA et al., 2020) e na discriminação de amostras de erva-mate de acordo com o país de origem (KAHMANN et al., 2017; MARCELO, POZEBON & FERRÃO, 2015, COZZOLINO, RESTANIO & FASIO, 2010), a similaridade das condições edafoclimáticas entre as regiões produtoras dentro do mesmo país aumenta o grau de similaridade entre os perfis químicos das amostras e contribui para a sobreposição entre as amostras que foi observada nesse estudo e em estudos similares (VIEIRA et al., 2020).

5.1.3 Avaliação da influência do tempo de armazenamento de amostras de erva-mate via PCA

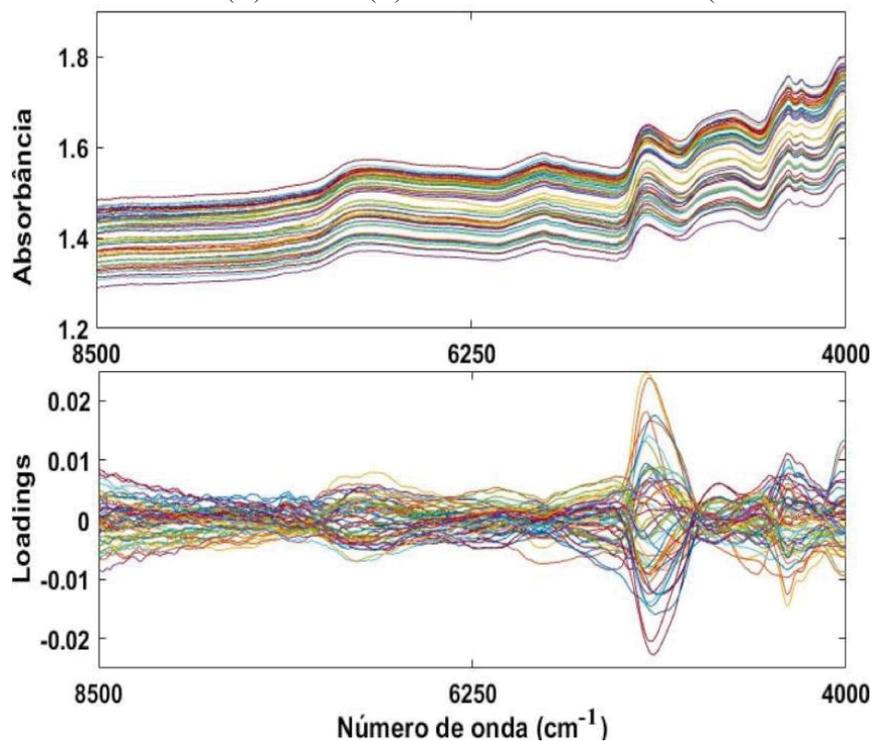
Matrizes alimentícias são perecíveis por natureza e sofrem modificações durante o processamento e o armazenamento devido à fatores intrínsecos (e.g. composição e disponibilidade de oxigênio) e extrínsecos (e.g. temperatura de armazenamento, tipo de embalagem, condições de umidade (RIBEIRO, 2007).

De acordo com as informações fornecidas pelos produtores, após aberta a erva-mate pode ser armazenada sob refrigeração por até 90 dias sem que suas características físico-químicas e sensoriais sejam alteradas de forma significativa.

Durante o período de isolamento social devido a pandemia de COVID-19 (~14 meses), as amostras de erva-mate (A01-A30) permaneceram armazenadas sob refrigeração. Entretanto, como o período de armazenamento foi extenso e as amostras estavam submetidas à umidade elevada, optou-se por coletar novamente os espectros (NIR com acessório DRIFT) das amostras a fim de avaliar a potencialidade da técnica de análise de componentes principais na discriminação das amostras de acordo com o tempo de armazenamento.

Os espectros foram coletados de acordo com o procedimento descrito no Item 3.3 desse documento. Nas avaliações preliminares a melhor discriminação das amostras de erva-mate de acordo com o período foi observada com a combinação de pré-processamentos (MSC+MC+SG) que corrigiu problemas com o espalhamento dos espectros e minimizou a presença de ruídos (FIGURA 36).

FIGURA 36 -ESPECTROS NIR DAS 30 AMOSTRAS DE ERVA-MATE COLETADOS EM DIFERENTES PERÍODOS ANTES (A) E APÓS (B) PRÉ-PROCESSAMENTO (MSC+MC+SG)

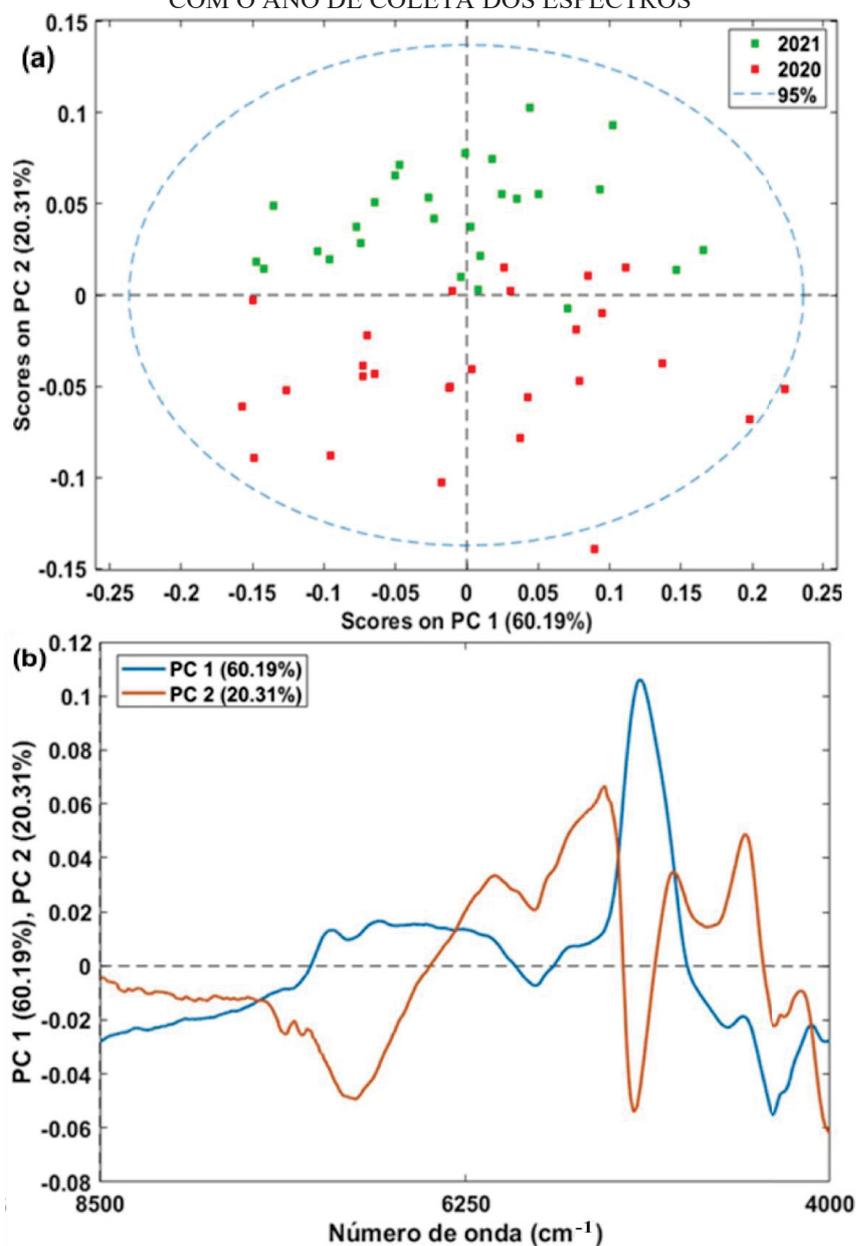


FONTE: A autora (2022).

Foram selecionadas cinco componentes principais para a aplicação da PCA, com uma variância acumulada de 98,5%. Na análise do gráfico de *scores* da PC1 versus PC2 (FIGURA 37a) pode ser observada a discriminação das amostras de erva-mate de acordo com o ano em que os espectros na região NIR foram coletados. É possível verificar que PC2 foi responsável pela discriminação das amostras, sendo que todas as amostras que tiveram o espectro coletado no ano de 2020 apresentam valores negativos em PC2 enquanto que as analisadas em 2021 apresentam valores positivos em PC2. No gráfico de *loadings* (FIGURA 37b) pode ser constatado que a discriminação das amostras em PC2 sofre maior influência de bandas de absorção em duas regiões distintas: entre 7000-6000 cm^{-1} e próximo a 5000 cm^{-1} . Essas bandas são associadas a estiramentos e deformações de ligações O-H, C-H, C-O que podem ser atribuídas ao aumento do teor de ~~umidade~~ e a modificações na composição química com relação aos teores de fenóis e lipídios (HANSEL et al., 2006).

A discriminação das amostras de acordo com o período de armazenamento, corrobora a hipótese de que para obter informações químicas representativas da realidade é necessária que seja estabelecida, dentro do protocolo de análise desse tipo de matriz, a janela de tempo entre a abertura da embalagem e a coleta dos espectros. Nesse estudo, o perfil espectral das amostras que estavam armazenadas sob refrigeração foi acompanhado diariamente e a janela para garantir a qualidade dos espectros foi definida em 1 semana após o preparo das amostras.

FIGURA 37 - GRÁFICOS DE *SCORES* (PC1 VS PC2) (A) E *LOADINGS* (B) OBTIDO A PARTIR DOS ESPECTROS NIR DAS AMOSTRAS COMERCIAIS DE ERVA-MATE IDENTIFICADAS DE ACORDO COM O ANO DE COLETA DOS ESPECTROS



FONTE: A autora (2022).

5.1.4 Identificação de amostras adoçadas de erva-mate via PLS-DA – região NIR

As evidências obtidas durante a etapa de análise exploratória, indicaram o potencial da espectroscopia na região NIR em fornecer informações relevantes para a discriminação de amostras com adição de sacarose. Entretanto, a fim de obter resultados mais robustos, foram construídos modelos de classificação utilizando o método supervisionado PLS-DA, que permite identificar as regiões com maior influência na separação das classes.

Idealmente, os modelos discriminatórios (e.g. PLS-DA), devem ser construídos a partir de classes balanceadas para evitar que exista viés em direção às classes majoritárias e declínio

da capacidade preditiva para as classes menos representadas (BALLABIO & CONSONNI, 2013). Entretanto, a disponibilidade de amostras adoçadas é limitada e por isso optou-se por coletar duplicatas das adoçadas e não incluir todas as não-adoçadas no modelo de classificação.

O banco de dados contendo os espectros NIR das amostras A31-A82 e as replicatas de algumas das amostras adoçadas, sendo a duplicata das amostras A77 e A80 e a triplicata da amostra A81 (totalizando 56 amostras), foi subdividido com o auxílio do algoritmo *Kennard-Stone* em dois conjuntos, o de calibração contendo 38 amostras (31 não-adoçadas e 7 adoçadas) e o de validação, contendo 18 amostras (12 não-adoçadas e 6 adoçadas). Vale ressaltar que o algoritmo de seleção foi aplicado de forma a manter as réplicas das amostras dentro do mesmo conjunto (calibração ou validação), evitando assim, o sobreajuste dos modelos construídos.

O melhor modelo foi obtido com os espectros NIR pré-processados por MSC+ Suavização SG, resultando nos menores erros de classificação e em valores de RMSECV e RMSEP de 0,1800 e 0,2845, respectivamente. O modelo foi construído com 5 variáveis latentes e apresentou uma variância acumulada de 98,54% em **X** e de 91,96% em **y** e atingiu 100% de classificações corretas no conjunto de calibração e apenas a amostra A42 do conjunto de validação que não foi avaliada com replicatas (indicada na FIGURA 38) foi classificada incorretamente como não-adoçada.

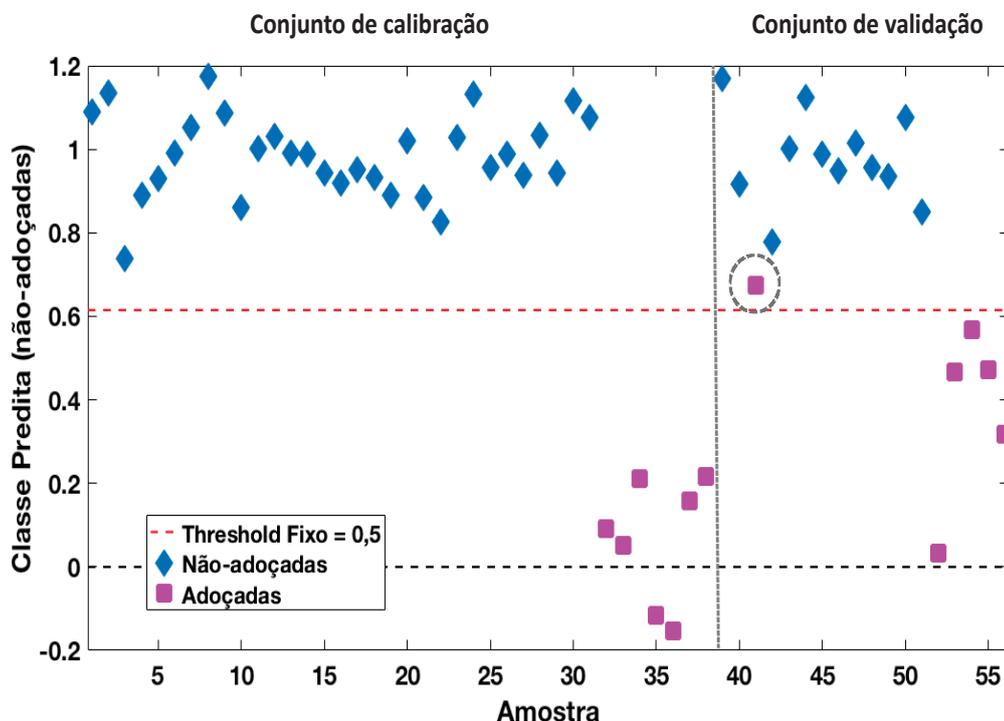
Como pode ser observado na TABELA 2, todas as amostras foram classificadas em uma das classes, e o modelo resultou em 83,3% de sensibilidade e especificidade para as amostras adoçadas e 100% de sensibilidade e especificidade para as amostras não-adoçadas.

TABELA 2 – Matriz de confusão do modelo de classificação PLS-DA construído com 5 VL's a partir dos dados da região NIR para a identificação de amostras de erva-mate adoçadas

	Calibração		Validação	
	Não-adoçadas	Adoçadas	Não-adoçadas	Adoçadas
Preditas como não-adoçada	31	0	12	1
Preditas como adoçada	0	7	0	5
Não atribuídas a nenhuma classe	0	0	0	0

FONTE: A autora (2023).

FIGURA 38 - RESULTADOS DO MODELO DE CLASSIFICAÇÃO PLS-DA CONSTRUÍDO COM 5 VL'S A PARTIR DOS DADOS DA REGIÃO NIR PARA A IDENTIFICAÇÃO DE AMOSTRAS DE ERVA-MATE ADOÇADAS

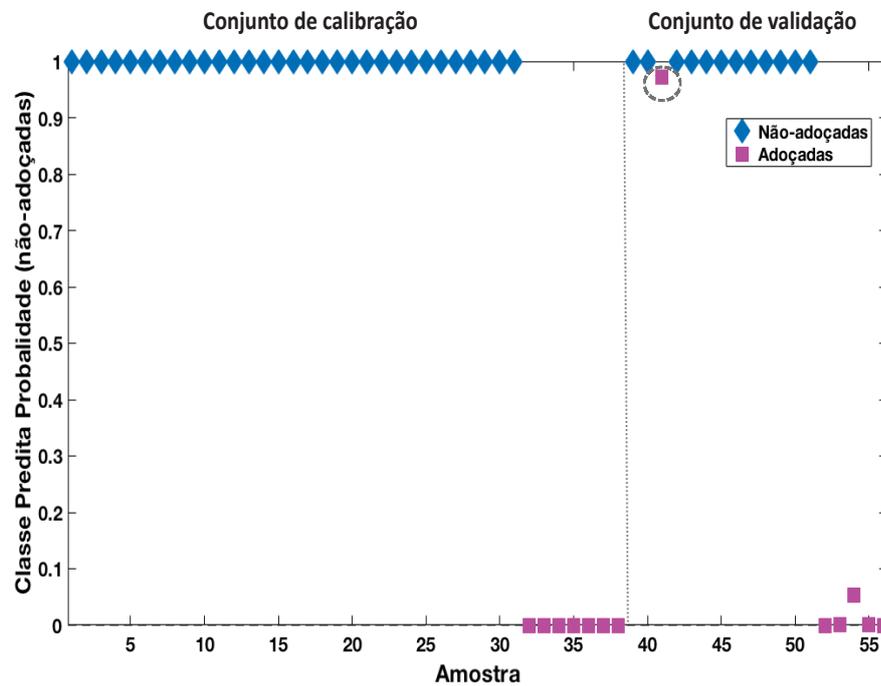


FONTE: A autora (2023).

Adicionalmente, no gráfico de probabilidade (FIGURA 39), é possível constatar que a amostra que foi classificada incorretamente possui acima de 90% de probabilidade de pertencer ao conjunto das amostras não-adoçadas. Uma vez que se tratam de amostras comerciais e que a legislação vigente não exige que o teor de açúcar seja declarado na embalagem, uma das possíveis causas da classificação incorreta pode ser o fato de que essa amostra possui um teor de sacarose significativamente menor do que as demais amostras adoçadas.

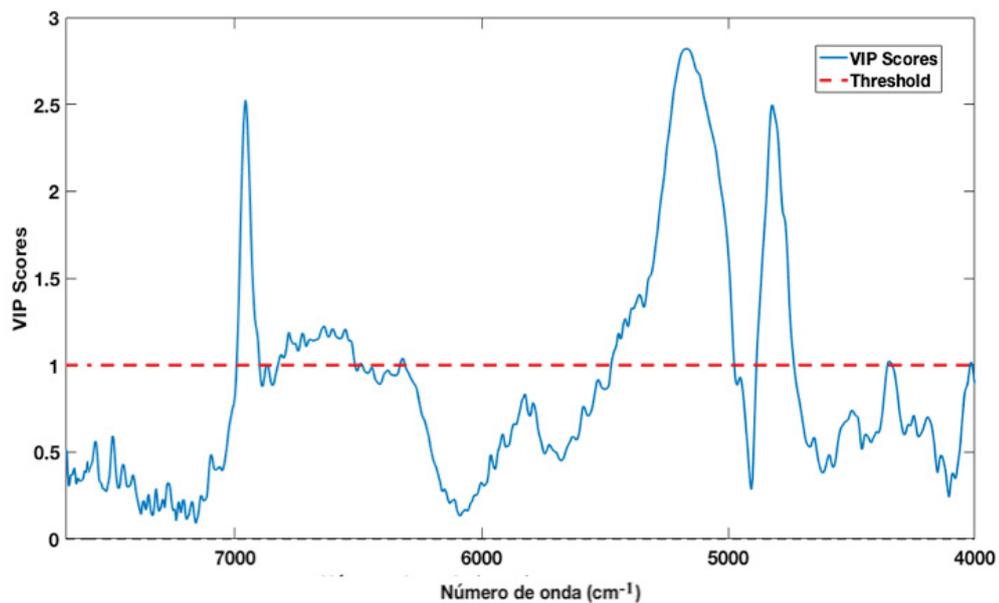
No gráfico de VIP (FIGURA 40), podem ser identificadas as regiões espectrais que mais contribuíram para a classificação das amostras. Como foi mencionado anteriormente, as bandas entre $7000\text{-}6000\text{ cm}^{-1}$ são atribuídas a compostos contendo múltiplas ligações O-H, como é o caso da sacarose. As duas regiões espectrais que se localizam entre $5300\text{-}4600\text{ cm}^{-1}$ são atribuídas a bandas de combinação de ligações O-H, C-O e C-H, provenientes de grupos fenólicos, amidos e lipídios e também ligações N-H, que podem estar relacionadas a presença de aminas primárias, provenientes das metilxantinas presentes na erva-mate (FRIZON et al., 2015; MAZUR et al., 2014; WORKMAN JR. & WEYER, 2012).

FIGURA 39 - VALORES DA PROBABILIDADE DE CLASSIFICAÇÃO DAS AMOSTRAS DE ERVA-MATE EM CADA UMA DAS CLASSES DO MODELO PLS-DA CONSTRUÍDO PARA A IDENTIFICAÇÃO DE AMOSTRAS DE ERVA-MATE ADOÇADAS



FONTE: A autora (2023).

FIGURA 40 - VIP SCORES DO MODELO PLS-DA CONSTRUÍDO COM 5 VL'S A PARTIR DOS DADOS DA REGIÃO NIR PARA A IDENTIFICAÇÃO DE AMOSTRAS DE ERVA-MATE ADOÇADAS



FONTE: A autora (2023).

Apesar do número limitado de amostras, o modelo desenvolvido no presente estudo apresentou resultados similares aos relatados na literatura. Vieira et al. (2020), construiu um modelo a partir de dados na região NIR utilizando 5 VL's e obteve valores de sensibilidade que variaram entre 80-97% nos conjuntos de calibração e que foram de 100% nos conjuntos de

predição. Lima (2019) conduziu um estudo comparativo ao de Vieira (2018) utilizando dados de Espectroscopia no Infravermelho Médio por Transformada de Fourier. Entretanto, além da necessidade de preparo das pastilhas com KBr, foi necessário o uso de 10 variáveis latentes para obtenção de valores de sensibilidade que variaram entre 62% e 75%, o que indica que as informações contidas na região espectral NIR apresentam maior relevância na discriminação de amostras adoçadas de erva-mate.

5.2 QUIMIOMETRIA APLICADA À DADOS DE ESPECTROSCOPIA DE RMN

5.2.1 Seleção do solvente deuterado

Conforme mencionado anteriormente nesse documento, o uso de solventes deuterados em análises de RMN contribui para a redução da sobreposição de sinais, facilitando a interpretação dos dados espectrais (NASCIMENTO, 2016). Sendo assim, na etapa de seleção do solvente a ser utilizado na obtenção dos espectros, os solventes deuterados água (D_2O) e metanol (CD_3OD) foram avaliados. A escolha desses solventes foi baseada na polaridade dos principais metabólitos presentes em amostras de erva-mate, dentre os quais se destacam, polifenóis, metilxantinas, flavonóides, saponinas e também aminoácidos, açúcares, minerais e vitaminas (CROGE, CUQUEL & PINTRO, 2020; WESTPHALEN et al., 2020).

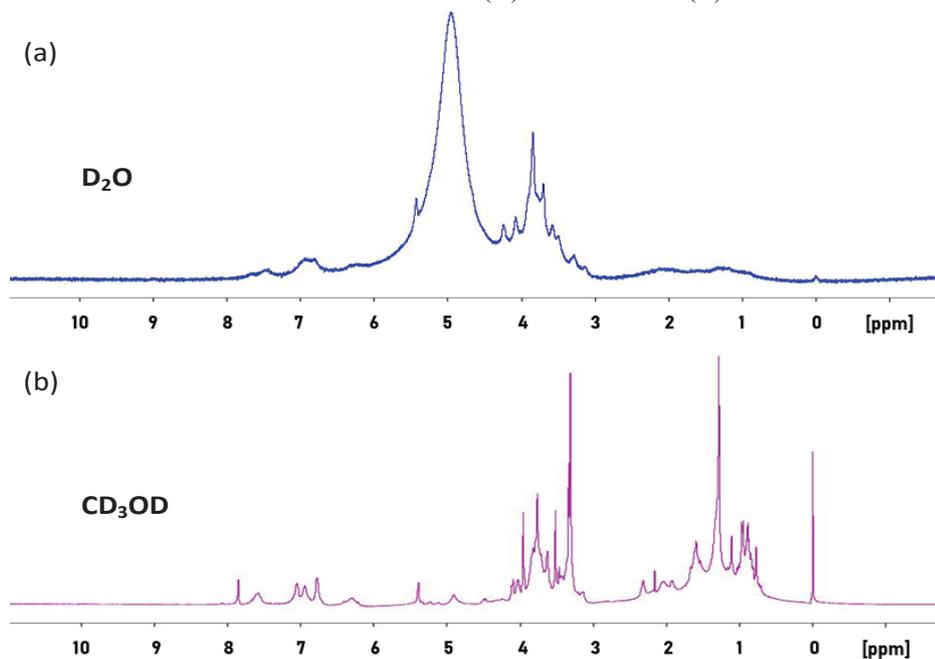
Na FIGURA 41 são apresentados os perfis metabólicos da erva-mate obtidos por RMN HR-MAS de 1H , e é possível constatar que a utilização do metanol deuterado permitiu a obtenção de um espectro com maior resolução, o que viabiliza a identificação de um maior número de metabólitos. Sendo assim, o CD_3OD foi escolhido para as análises.

5.2.2 Análise do perfil metabólico obtido via RMN HR-MAS de 1H

O perfil metabólico médio obtido por RMN HR-MAS de 1H das 50 amostras de erva-mate (B01-B50), é apresentado na FIGURA 42. Os espectros apresentaram alto nível de sobreposição de sinais, não sendo possível observar diferenças entre as amostras de diferentes regiões, o que corrobora a necessidade de aplicação de estratégias de análise multivariadas para a identificação de padrões característicos que possam auxiliar na discriminação.

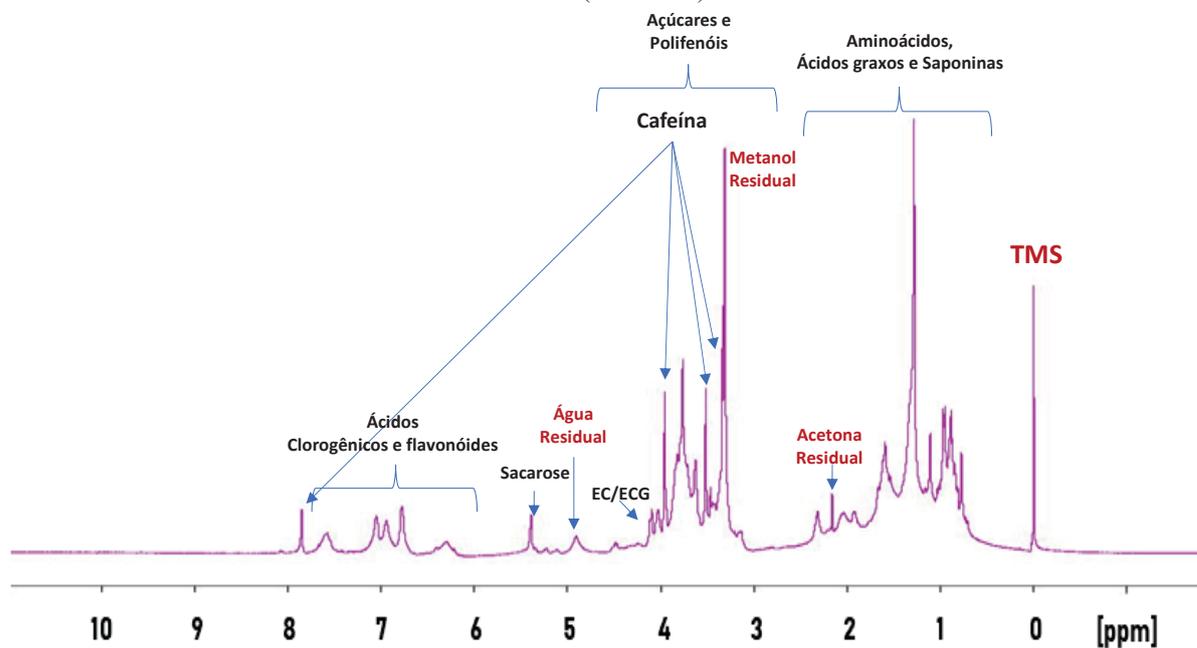
Na análise do espectro de RMN HR-MAS de 1H (FIGURA 42), foi possível observar os deslocamentos químicos característicos do padrão TMS (δ_H 0,00) e dos sinais residuais de água (δ_H 4,80-4,70), metanol (δ_H 3,29-3,31) e da acetona (δ_H 2,16-2,14) que foi utilizada como solvente para limpeza do rotor entre as amostras (SANTOS, 2023).

FIGURA 41 - ESPECTROS DE RMN HR-MAS DE ^1H (400 MHz) DE AMOSTRAS COMERCIAIS DE ERVA-MATE OBTIDOS COM ÁGUA (A) E METANOL (B) DEUTERADOS



FONTE: A autora (2023).

FIGURA 42 – PERFIL METABÓLICO MÉDIO DE AMOSTRAS COMERCIAIS DE ERVA-MATE OBTIDOS VIA RMN HR-MAS DE ^1H (400 MHz) EM METANOL DEUTERADO



FONTE: SANTOS (2023); TAVARES & FERREIRA, 2006; (SANTOS, ALVAREGA & BOFFO, 2020).

Foi possível observar a presença dos quatro simpletos característicos da cafeína, sendo três sinais atribuídos à grupamentos metila (δ_{H} 3,33, 3,50 e 3,93) e um sinal característico do hidrogênio aromático (δ_{H} 7,78). Além disso, os sinais na região aromática entre δ_{H} 7,76-6,50

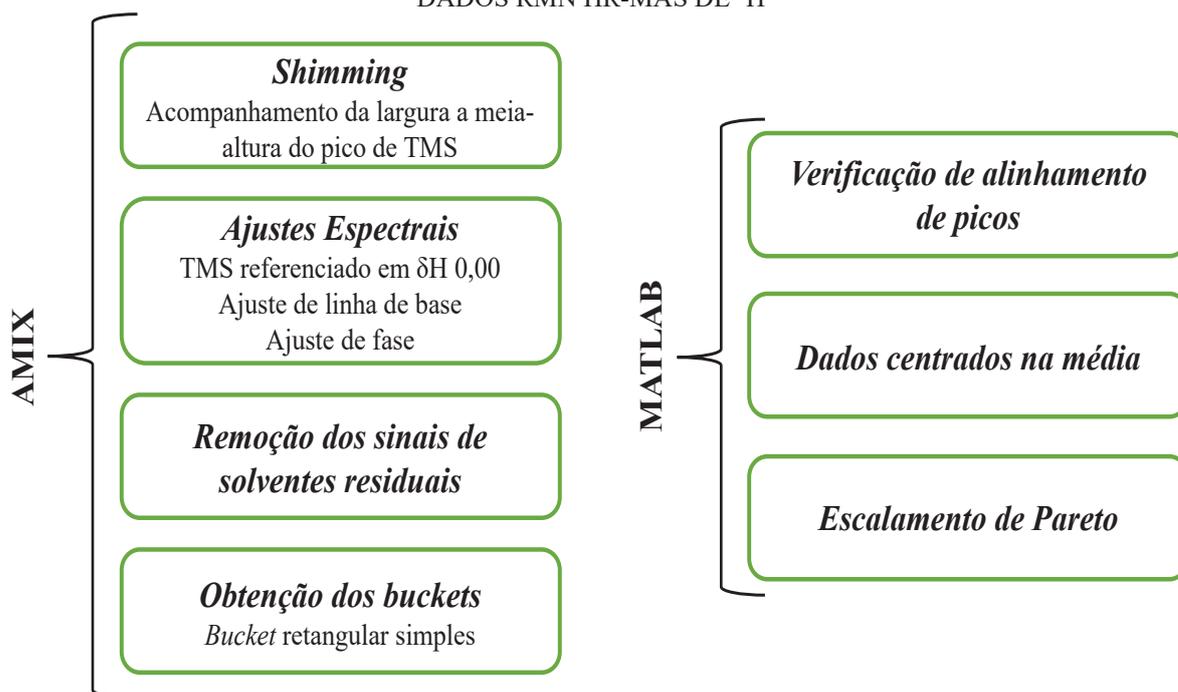
ppm podem ser atribuídos à presença de ácidos clorogênicos e à prótons presentes nos anéis aromáticos de flavonóides (TAVARES & FERREIRA, 2006).

Os espectros de RMN ^1H também fornecem sinais em deslocamentos químicos que podem ser associados à presença de aminoácidos e saponinas específicas do gênero *Ilex* (e.g. prótons alifáticos das moléculas de leucina, valina e matesaponinas (δ_{H} 1,1 a 0,9), polifenóis (e.g. (-)-epi-catequina (EC) e (-)-epi-galocatequina δ_{H} 4,2 a 4,1)), carboidratos (e.g. sinais típicos de hidrogênios de carboidratos δ_{H} 4,0 a 3,2 além do hidrogênio anomérico presente na molécula de sacarose e apresentado como um duplete em δ_{H} 5,4) e ácidos graxos de cadeia longa encontrados em diferentes tipos de chás (δ_{H} 1,32 a 1,28) (SANTOS, ALVAREGA & BOFFO, 2020; MENG et al., 2017; PRESTES et al., 2012).

5.2.3 Análise exploratória via PCA – RMN HR-MAS de ^1H

A fim de garantir que as variações observadas nos espectros estivessem relacionadas apenas com as diferenças de composição das amostras, foi necessária a aplicação de uma série de pré-processamentos espectrais dentro dos softwares TopSpin - AMIX e MATLAB. As etapas de pré-processamento aplicadas aos dados são demonstradas de maneira simplificada no fluxograma da FIGURA 43.

FIGURA 43 - ESQUEMA SIMPLIFICADO DAS ETAPAS DE PRÉ-PROCESSAMENTO APLICADAS AOS DADOS RMN HR-MAS DE ^1H



FONTE: A autora (2023).

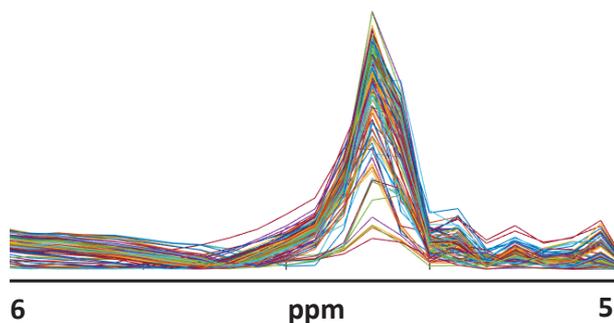
A primeira etapa consistiu no ajuste da homogeneidade do campo magnético principal (B_0), conhecida como *shimming*. Esse processo é uma etapa crucial na obtenção de espectros com maior resolução espectral, pois na presença de um campo magnético não-uniforme podem ocorrer frequências de ressonância ligeiramente diferentes, o que resultará em amplas linhas espectrais (WONG & LUCAS-TORRES, 2018). Segundo Santos (2019), a etapa de *shimming* consiste em garantir a reprodutibilidade do posicionamento do rotor e do volume de solvente na sonda para assegurar que as variações de frequência sejam menores que 1 Hz. Sendo assim, a qualidade da etapa de *shimming* foi monitorada através da largura a meia altura do sinal do TMS (FIGURA 48), tendo sido obtidos valores abaixo de 1 Hz em todos os espectros.

Na segunda etapa, foram efetuados ajustes de linha de base e da fase dos espectros. Além disso, os espectros foram referenciados em relação ao sinal do TMS (δ_H 0,00) e uma seleção de variáveis *a priori* foi efetuada a partir da remoção das regiões espectrais contendo sinais residuais de água (δ_H 4,80-4,70), metanol (δ_H 3,31-3,29) e da acetona (δ_H 2,16-2,14) que foi utilizada como solvente para limpeza do rotor entre as amostras.

A terceira etapa consistiu na geração dos *buckets* com foco na obtenção de um conjunto de dados formado por regiões espectrais que contemplassem a variabilidade dos sinais. Nessa etapa, foi utilizado o método de *buckets* retangulares simples, que consiste na geração de regiões de mesmo tamanho de acordo com o valor atribuído pelo analista (WANG, MALDONADO-DEVINCCI & JIANG, 2020). Os tamanhos de *bucket* avaliados variaram de 0,01 a 0,1 ppm.

Após a geração dos *buckets*, os dados foram exportados para ambiente MATLAB onde os dados foram avaliados com relação ao alinhamento dos picos. Como pode ser observado na FIGURA 44, os espectros não apresentaram problemas relacionados ao alinhamento de picos e então foi efetuada a segunda etapa de pré-processamentos que se baseou em centrar os dados na média e autoescalar pelo método de Pareto a fim de garantir que diferenças de intensidade não afetassem a interpretação dos resultados.

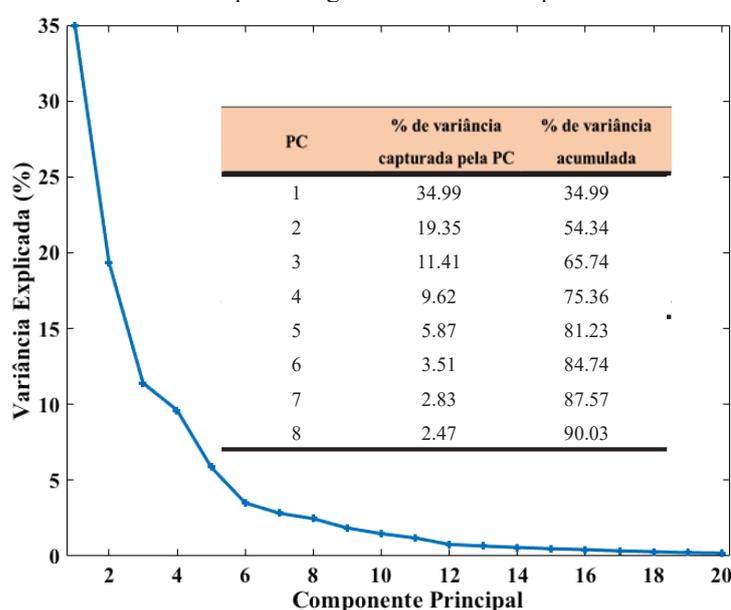
FIGURA 44 – DESTAQUE DE REGIÃO ESPECTRAL ENTRE 6,0-5,0 PPM NO *BUCKET* DE 0,05 PPM



FONTE: A autora (2023).

Todos os *buckets* gerados foram avaliados de acordo com potencialidade de discriminação das amostras de acordo com a origem geográfica, sendo que as melhores tendências de discriminação foram obtidas com o *bucket* de 0,05 ppm. A relação entre as componentes principais (PC's) e a porcentagem de variância explicada (FIGURA 45) mostra que as 8 primeiras PC's acumularam mais de 90% da variabilidade total existente nos dados. Como a adição de um maior número de PC's não resultou em um melhor agrupamento das amostras em relação aos parâmetros avaliados, apenas essas 8 foram avaliadas.

FIGURA 45 - VARIÂNCIA EXPLICADA *VERSUS* DO NÚMERO DE PC'S PARA DADOS DE RMN ¹H.
INSERIDO: tabela de porcentagem de variância capturada em cada PC

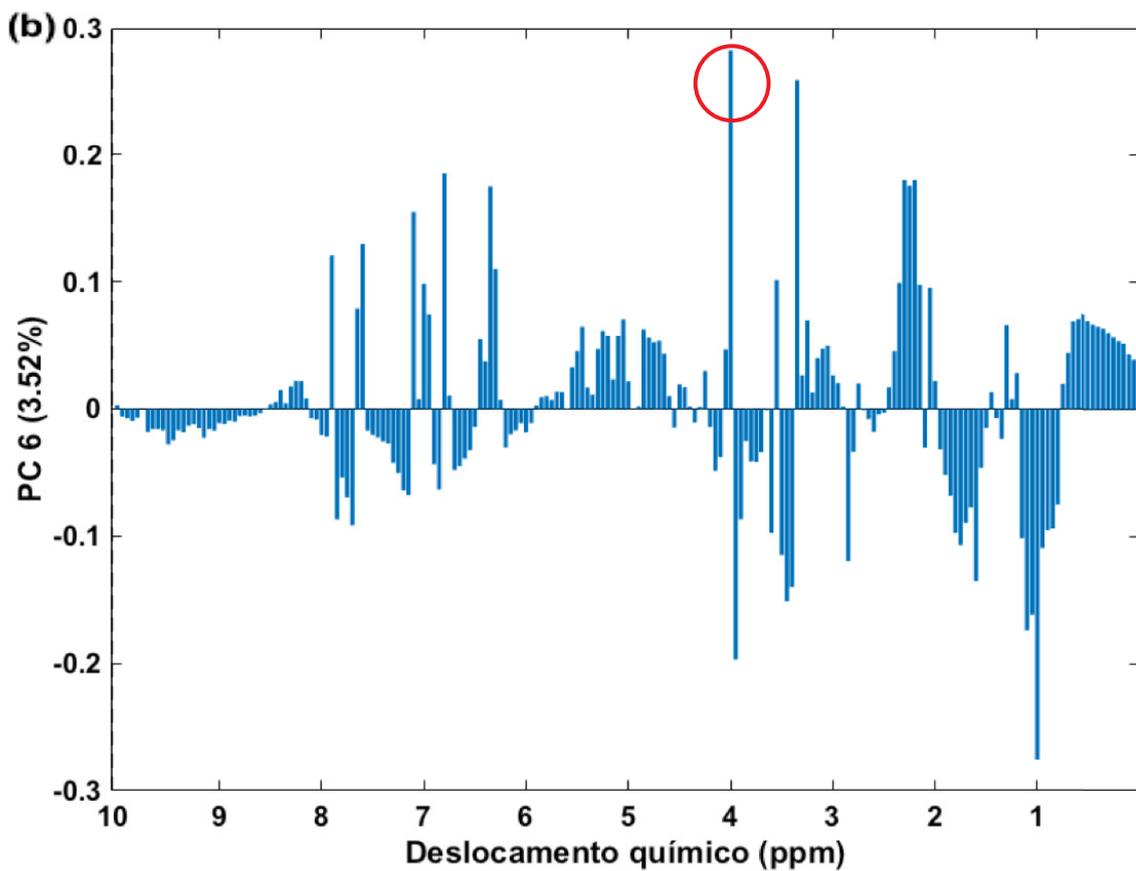
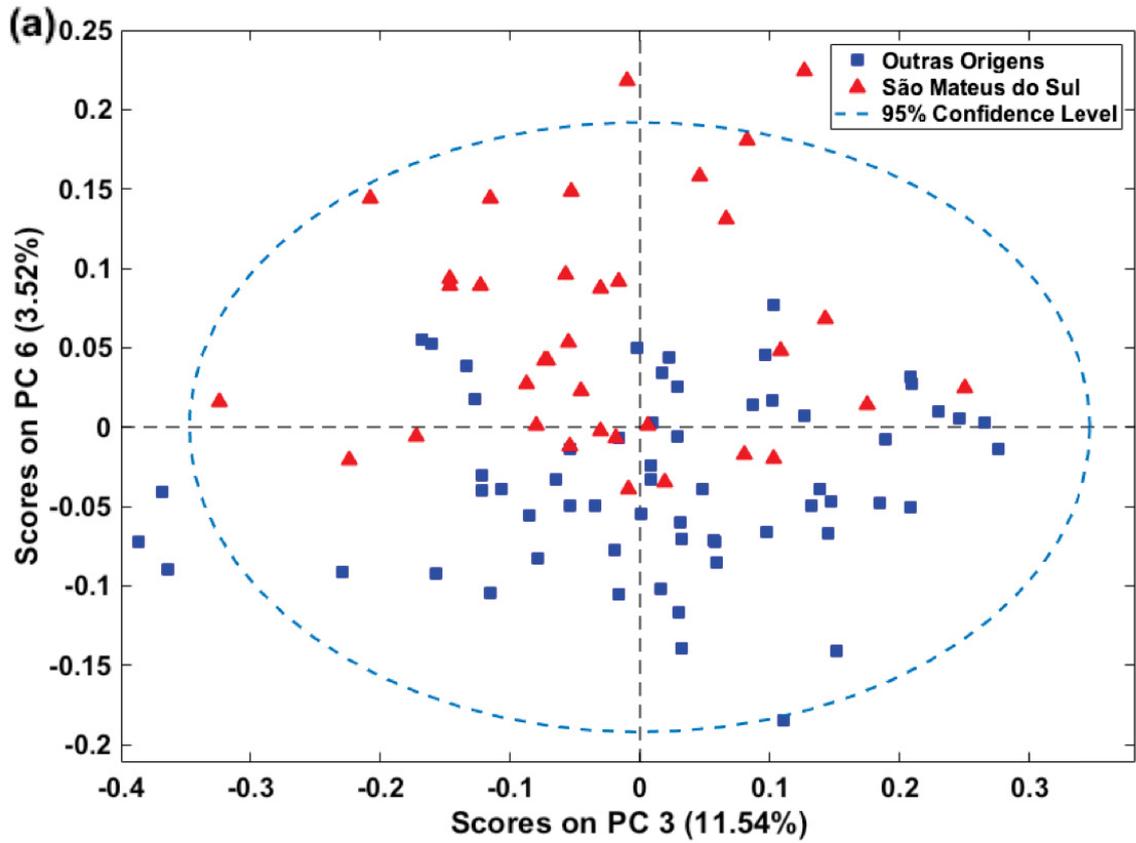


FONTE: A autora (2023).

Visando a melhor visualização nos gráficos de *scores*, as amostras foram identificadas com cores distintas de acordo com a região de produção: sendo São Mateus do Sul (vermelho), e todas as outras origens geográficas (azul). Após a seleção do número de PC's, todas as possíveis combinações das mesmas foram avaliadas em busca de tendências de separação e agrupamento das amostras de erva-mate de acordo com a origem geográfica. Entretanto, a combinação de PC3 e PC6 foi a que apresentou as melhores tendências de separação.

No gráfico de *scores* de PC3 *versus* PC6 (FIGURA 46a) pode ser observado que as amostras de São Mateus do Sul se encontram deslocadas para o lado positivo de PC6, enquanto as das demais localidades apresentam *scores* negativos nessa PC. No gráfico de *loadings* para PC6 (FIGURA 46b) é possível constatar que as amostras localizadas no lado negativo sofrem influência principalmente dos *buckets* na região de prótons aromáticos (δ_H 8,00-6,00) e na região característica da presença de sinais da cafeína (δ_H 4,00-3,00).

FIGURA 46 - GRÁFICOS DE SCORES (PC3 VS PC6) (A) E LOADINGS (B) OBTIDO A PARTIR DOS ESPECTROS DE RMN ^1H DAS AMOSTRAS COMERCIAIS DE ERVA-MATE IDENTIFICADAS DE ACORDO COM A ORIGEM GEOGRÁFICA



FONTE: A autora (2023).

Esses resultados complementam os obtidos a partir da espectroscopia na região do infravermelho, na qual foi possível verificar que as amostras de São Mateus do Sul apresentam maiores valores de *scores* na região caracterizada pela presença de compostos fenólicos e flavonóides, o que pode ser correlacionado com a produção da erva em áreas sombreadas (BRAGHINI et al., 2014; CRAVO FILHO et al., 2017; DA SILVEIRA et al., 2016).

É interessante observar que apesar de serem conhecidas pelo sabor mais suave e adocicado, as amostras de São Mateus do Sul apresentaram sinais mais intensos na região característica da presença da polifenóis e cafeína (circulado na Figura 46B) que são compostos associados à sabores mais intensos e amargos (POOLE & TORDOFF, 2017; PAGLIOSA et al., 2009). Porém, os sabores são resultado de processos complexos que envolvem a interação entre compostos químicos durante o processamento industrial, além de ser a resposta da percepção subjetiva do consumidor que pode ser influenciado por fatores relacionados às expectativas e preferências (YANG & LEE, 2020; SANTAGIULIANA et al., 2019; ASHTON, 2014).

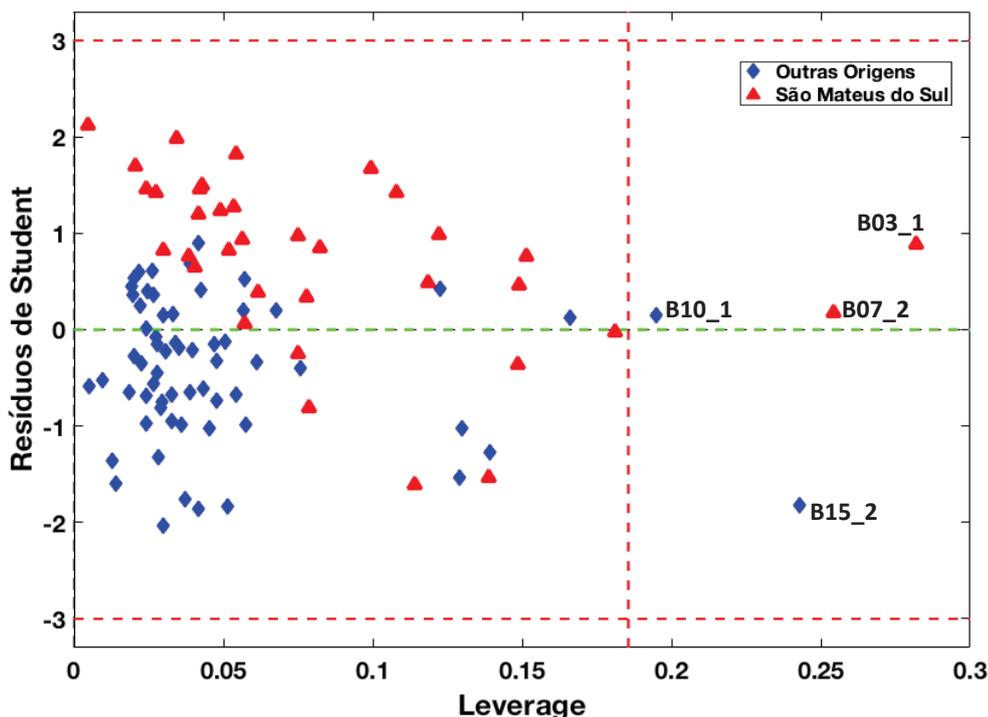
Os *buckets* da região caracterizada pela presença de aminoácidos, ácidos graxos e saponinas (δ_H abaixo de 1,00 e entre 2,50-2,00) também influenciaram a separação das amostras para o lado positivo de PC6, indicando que existem diferenças significativas no perfil metabólico das amostras produzidas em São Mateus com relação à essas classes de compostos.

5.2.4 Classificação das amostras de erva-mate de acordo com a origem geográfica via PLS-DA – região de RMN

As evidências obtidas durante a etapa de análise exploratória, indicaram o potencial da espectroscopia de RMN 1H em fornecer informações relevantes para a discriminação das amostras comerciais de erva-mate de acordo com a origem geográfica. Sendo assim, a fim de obter resultados mais robustos, foram construídos modelos de classificação utilizando o método de reconhecimento de padrões supervisionado, PLS-DA, que permite identificar as regiões com maior influência na separação das classes.

A detecção de amostras com comportamento anômalo foi efetuada durante a etapa de otimização dos modelos de calibração através da avaliação dos valores de resíduos de *Student* e *leverage* (FIGURA 47). Quatro amostras (B03_1, B07_2, B10_1 e B15_2) apresentaram valores de *leverage* acima do limite, o que indica que as amostras apresentam um perfil que se distancia significativamente da média do conjunto. O comportamento anômalo pode ter sido causado por erros no preparo da replicatas ou interferências durante a coleta dos espectros, sendo assim, essas amostras foram removidas antes que as métricas de desempenho do modelo de classificação fossem avaliadas.

FIGURA 47 - RESÍDUOS DE *STUDENT* VS. *LEVERAGE*: MODELO PLS-DA PARA CLASSIFICAÇÃO DE AMOSTRAS COMERCIAIS DE ERVA-MATE DE ACORDO COM A ORIGEM GEOGRÁFICA



FONTE: A autora (2023).

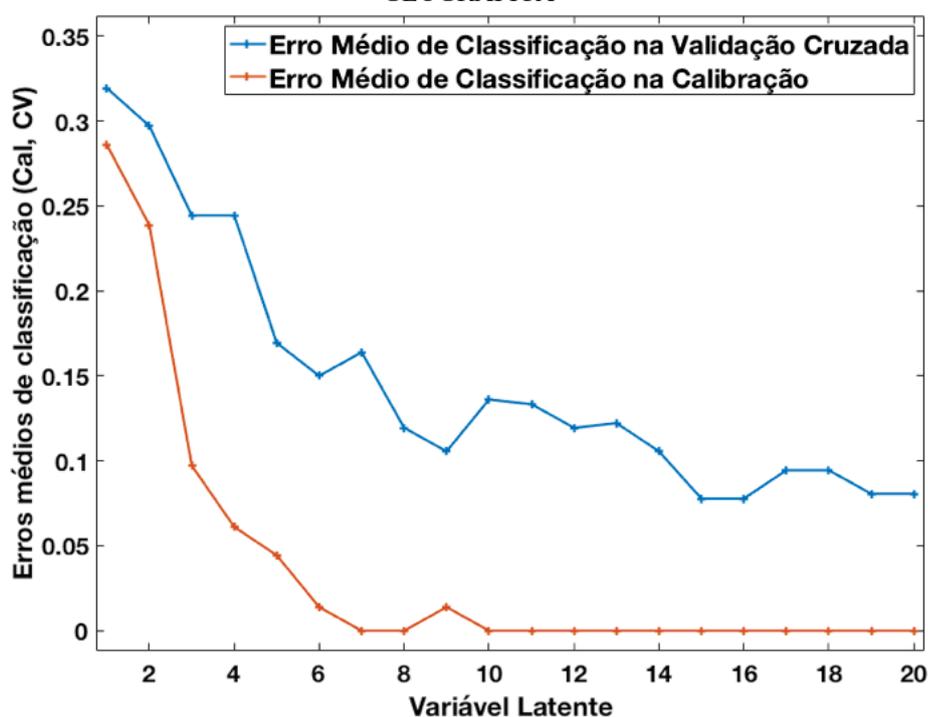
O banco de dados contendo os espectros de RMN ^1H das amostras B01-B50 em duplicata (com exceção das amostras B03, B07, B10 e B15), totalizando 96 amostras, foi subdividido com o auxílio do algoritmo *Kennard-Stone* em dois conjuntos, o de calibração contendo 66 amostras (30 de São Mateus do Sul e 36 de Outras Origens) e o de validação, contendo 30 amostras (7 de São Mateus do Sul e 23 de Outras Origens). Vale ressaltar que o algoritmo de seleção foi aplicado de forma a manter as réplicas das amostras dentro do mesmo conjunto (calibração ou validação), evitando assim, o sobreajuste dos modelos construídos.

O melhor modelo foi obtido com os espectros de RMN ^1H pré-processados por MC+Escalamento de Pareto e foi construído com 6 variáveis latentes que apresentaram uma variância acumulada de 59,56% em \mathbf{X} e de 80,51% em \mathbf{y} e resultaram na redução dos valores de erro de predição tanto para a calibração quanto para a validação cruzada (FIGURA 48).

Para o conjunto de calibração, o modelo resultou em 83,3% de classificações corretas das amostras de Outras Origens e 86,7% das amostras de São Mateus do Sul. Pode ser observado na FIGURA 49 que, no conjunto de calibração, com exceção da amostra B18 que teve as duas replicatas classificadas como sendo originária de São Mateus, as outras amostras (B08, B17, B23, B26, B28, B37) tiveram apenas uma das réplicas classificadas incorretamente, o que pode ser entendido como um indicativo de problemas relacionados à coleta dos espectros e não à uma falha no resultado de classe prevista pelo modelo PLS-DA.

Apesar da amostra B18 ser proveniente do Estado de Santa Catarina, a cidade de Canoinhas tem proximidade geográfica com São Mateus do Sul, o que pode justificar a similaridade de perfil metabólico entre essas amostras e o conjunto de São Mateus. Adicionalmente, foi observado que as amostras B17, B18 e B37 são comercializadas como provenientes do cultivo orgânico, fato que pode ter influenciado a diferenciação dessas amostras em relação ao conjunto original e gerado o erro de classificação.

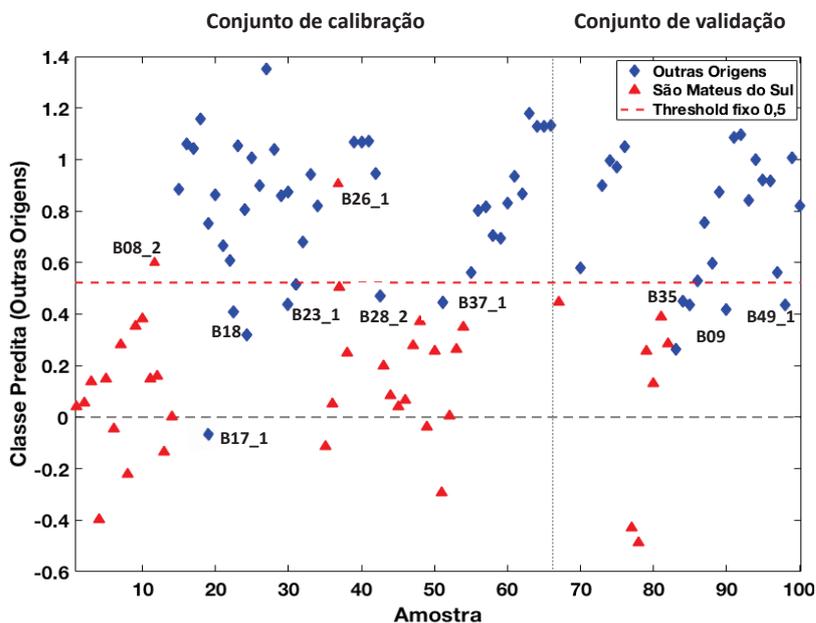
FIGURA 48 - ERROS DE CLASSIFICAÇÃO VS VARIÁVEIS LATENTES: MODELO PLS-DA PARA CLASSIFICAÇÃO DE AMOSTRAS COMERCIAIS DE ERVA-MATE DE ACORDO COM A ORIGEM GEOGRÁFICA



FONTE: A autora (2023).

Com relação aos conjuntos de validação, pode ser observado na FIGURA 49 e na TABELA 3 que todas as amostras foram classificadas em uma das classes, e o modelo resultou em 78,3% de classificações corretas das amostras de outras origens e de 100% de classificações corretas para as de São Mateus do Sul. Foi possível constatar que as duas replicatas das amostras B09 e B35, além de uma replicata da amostra B49 (produzidas em SC ou RS) foram classificadas como pertencentes ao conjunto de São Mateus, o que indica que a amostra apresenta um perfil metabólico mais similar ao dessa região. Entretanto, não foi possível identificar correlações entre as amostras baseadas nas informações fornecidas nos rótulos.

FIGURA 49 - RESULTADOS DO MODELO DE CLASSIFICAÇÃO PLS-DA CONSTRUÍDO COM 6 VL'S A PARTIR DOS DADOS DE RMN ¹H PARA A DISCRIMINAÇÃO DE AMOSTRAS DE ERVA-MATE DE ACORDO COM A ORIGEM GEOGRÁFICA



FONTE: A autora (2023).

TABELA 3 – Matriz de confusão do modelo de classificação PLS-DA construído com 6 VL's a partir dos dados de RMN ¹H para a discriminação de amostras de erva-mate de acordo com a origem geográfica

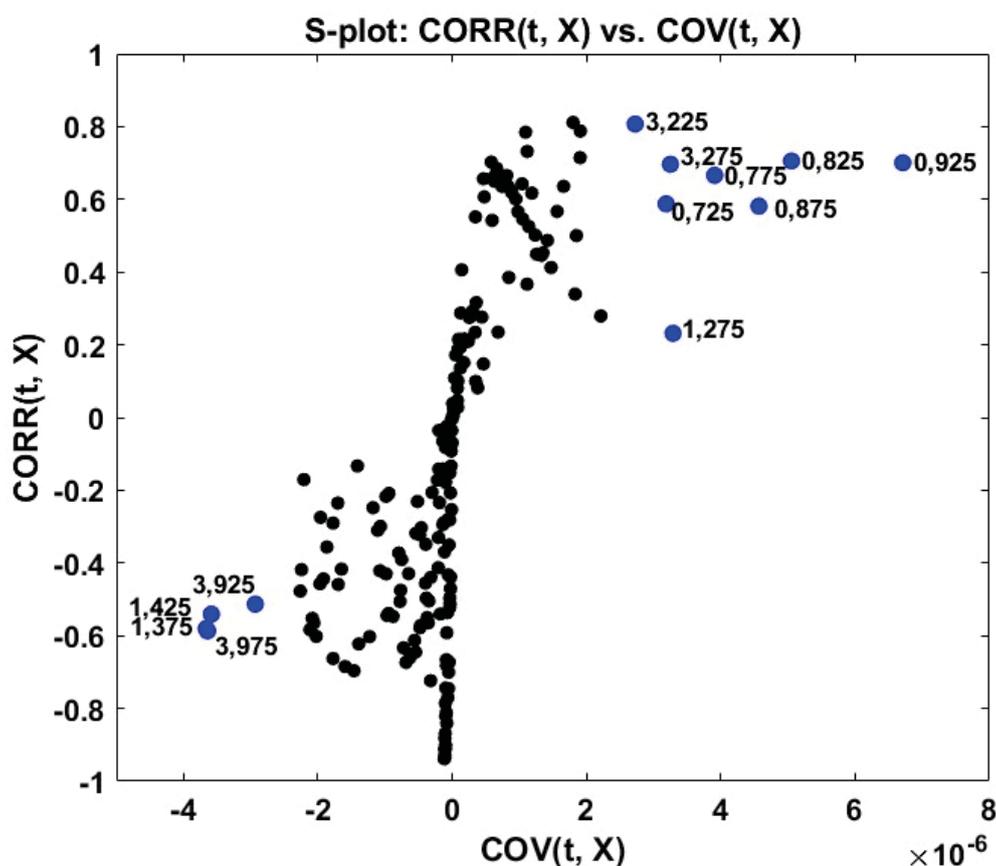
	Calibração		Validação	
	Outras Origens	São Mateus do Sul	Outras Origens	São Mateus do Sul
Preditas como Outras origens	30	4	18	0
Preditas como São Mateus do Sul	6	26	5	7
Não atribuídas a nenhuma classe	0	0	0	0

FONTE: A autora (2023).

A identificação das variáveis que mais contribuíram para a discriminação das amostras de acordo com a origem geográfica foi efetuada a partir da análise do S-plot (FIGURA 50). O S-plot combina a covariância modelada (eixo x) e a correlação modelada (eixo y) em um gráfico de dispersão, permitindo a identificação dos sinais que apresentam a maior influência no modelo (SHYTI et al., 2015). Os resultados do S-plot indicam a influência significativa dos *buckets* com deslocamentos químicos entre δ_H 4,00-3,20 que possui sinais característicos de cafeína, açúcares e polifenóis e também na região entre δ_H 1,5-0,7 caracterizada por deslocamentos que podem ser atribuídos à aminoácidos, ácidos graxos e saponinas.

Conforme mencionado anteriormente, o perfil metabólico da erva-mate é significativamente dependente de diversos fatores, que incluem desde as diferentes estratégias de cultivo produção até o tipo de embalagem. Sendo assim, a elucidação dos compostos específicos que influenciam a discriminação das amostras de acordo com a origem geográfica foi obtida a partir de análise com as técnicas de RMN bidimensionais (COSY, HSQC e HMBC).

FIGURA 50 – S-PLOT DO MODELO PLS-DA CONSTRUÍDO COM 6 VL'S A PARTIR DOS DADOS DE RMN ^1H PARA A DISCRIMINAÇÃO DE AMOSTRAS DE ERVA-MATE DE ACORDO COM A ORIGEM GEOGRÁFICA



FONTE: A autora (2023).

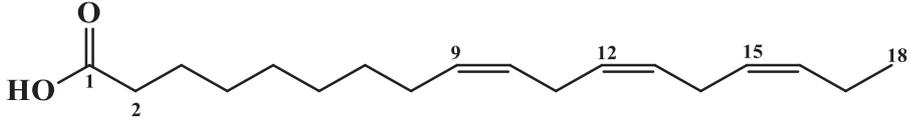
5.2.5 Elucidação estrutural de marcadores via RMN bidimensional

5.2.5.1 Identificação estrutural do ácido linolênico (1)

A análise de RMN HR-MAS de ^1H permitiu observar a presença de um tripleto em δ_{H} 0,97 (t , $J = 7,5$, 3H) que apresentou correlação direta no mapa de correlação (^1H - ^{13}C) HSQC com o carbono em δ_{C} 15,1 (C18), característico de um grupamento metílico. A correta posição desse grupo foi estabelecida através do mapa de correlação a longa distância (^1H - ^{13}C) HMBC, no qual o sinal em 0,97 ppm correlacionou a longa distância com os carbonos em δ_{C} 21,6 (C17)

e 129,4 (C16). Verificou-se também, através do mapa de correlação direta (^1H - ^{13}C) HSQC, a presença de um sinal em δ_{H} 2,32 (*t*, $J = 7,6$, 2H), que está ligado ao carbono em δ_{C} 35,1 (C2). Além disso, pôde-se observar que o sinal em 2,31 ppm correlacionou a longa distância com os carbonos em δ_{C} 26,2 (C3), 30,5 (C4) e 174,7 (C1). De acordo com os dados de RMN HR-MAS de ^1H e RMN 2D, bem como comparação com a literatura foi possível confirmar a presença do ácido linolênico (FIGURAS 51 a 53; TABELA 4) (ALAMSJAH *et al.*, 2005).

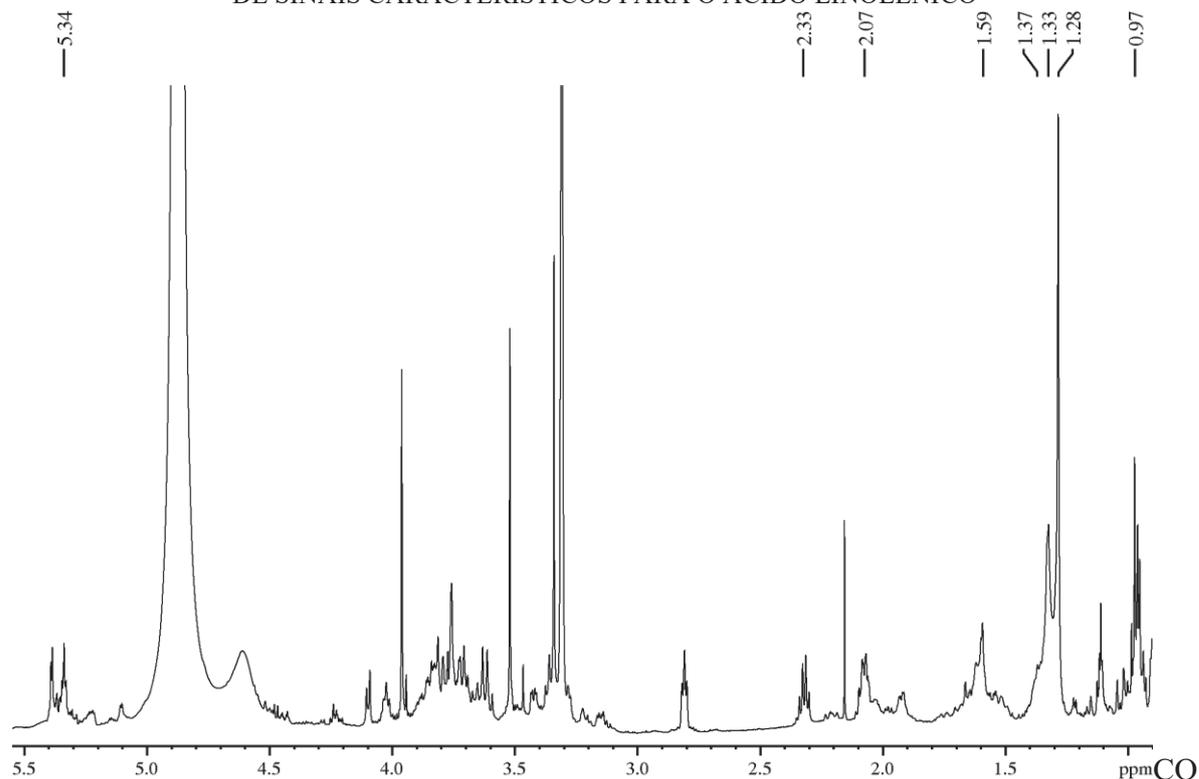
TABELA 2 - DADOS DE RMN HR-MAS DE ^1H E RMN 2D PARA O ÁCIDO LINOLÊNICO



Posição	Metabólito identificado			Literatura	
	$\delta_{\text{H}}^{\text{a,b}}$	HSQC ^b	HMBC ^b	$\delta_{\text{H}}^{\text{c}}$	$\delta_{\text{C}}^{\text{c}}$
1	-	174,7	-	-	180,6
2	2,32 (<i>t</i> , $J = 7,6$, 2H)	35,1	C1/C3/C4	2,35 (<i>t</i> , $J = 7,2$)	34,0
3	1,59 (<i>m</i> , 2H)	26,2	C4	-	29,1
4	128-1,32 (<i>m</i> , 2H)	30,5	C4	-	-
5	2,07 (<i>m</i> , 2H)	28,3	C4	-	-
6	-	-	-	-	-
7	-	-	-	-	-
8	1,37 (<i>m</i> , 2H)	-	-	-	25,6
9	-	-	-	-	127,6
10	5,34 (<i>m</i> , 1H)	129,4	C3	-	128,1
11	-	-	-	-	27,2
12	5,34 (<i>m</i> , 1H)	129,4	C3	5,28-5,43 (<i>m</i>)	130,1
13	-	-	-	-	-
14	-	-	-	-	29,9
15	-	-	-	5,28-5,43 (<i>m</i>)	128,2
16	5,34 (<i>m</i> , 1H)	129,4	C3	5,28-5,43 (<i>m</i>)	131,9
17	-	21,6	-	-	20,6
18	0,97 (<i>t</i> , $J = 7,5$, 3H)	15,1	C16/C17	0,98 (<i>t</i> , $J = 7,6$)	14,3

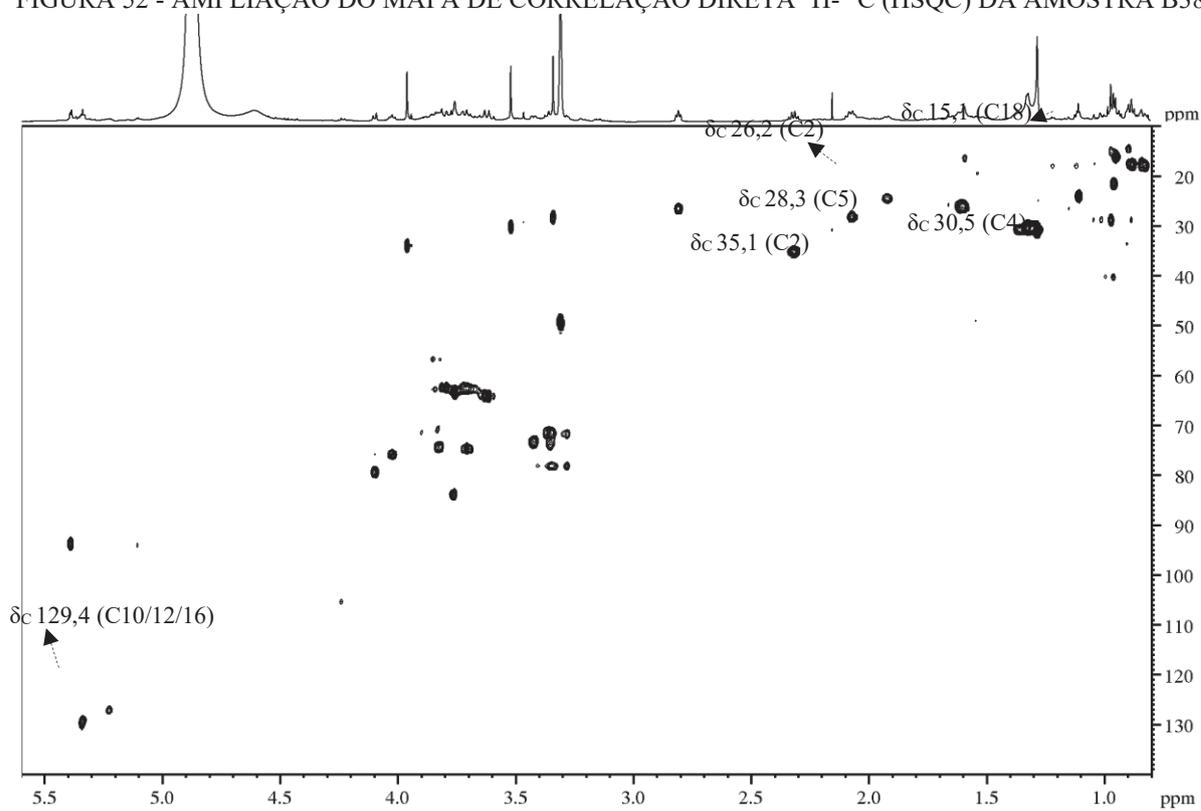
LEGENDA: ^AEXPERIMENTO REALIZADO A 400 MHZ PARA ^1H EM CD_3OD . ^BEXPERIMENTO REALIZADO A 600 MHZ PARA ^1H E 150 MHZ PARA ^{13}C EM CD_3OD . ^{A,B}TMS COMO PADRÃO DE REFERÊNCIA. ^CALAMSJAH E COLABORADORES (2005) (400 MHZ PARA ^1H , CDCl_3/TMS).

FIGURA 51 - ESPECTRO DE RMN DE ^1H (δ_{H} 0,8-5,60) DA AMOSTRA B38 INDICANDO A PRESENÇA DE SINAIS CARACTERÍSTICOS PARA O ÁCIDO LINOLÊNICO



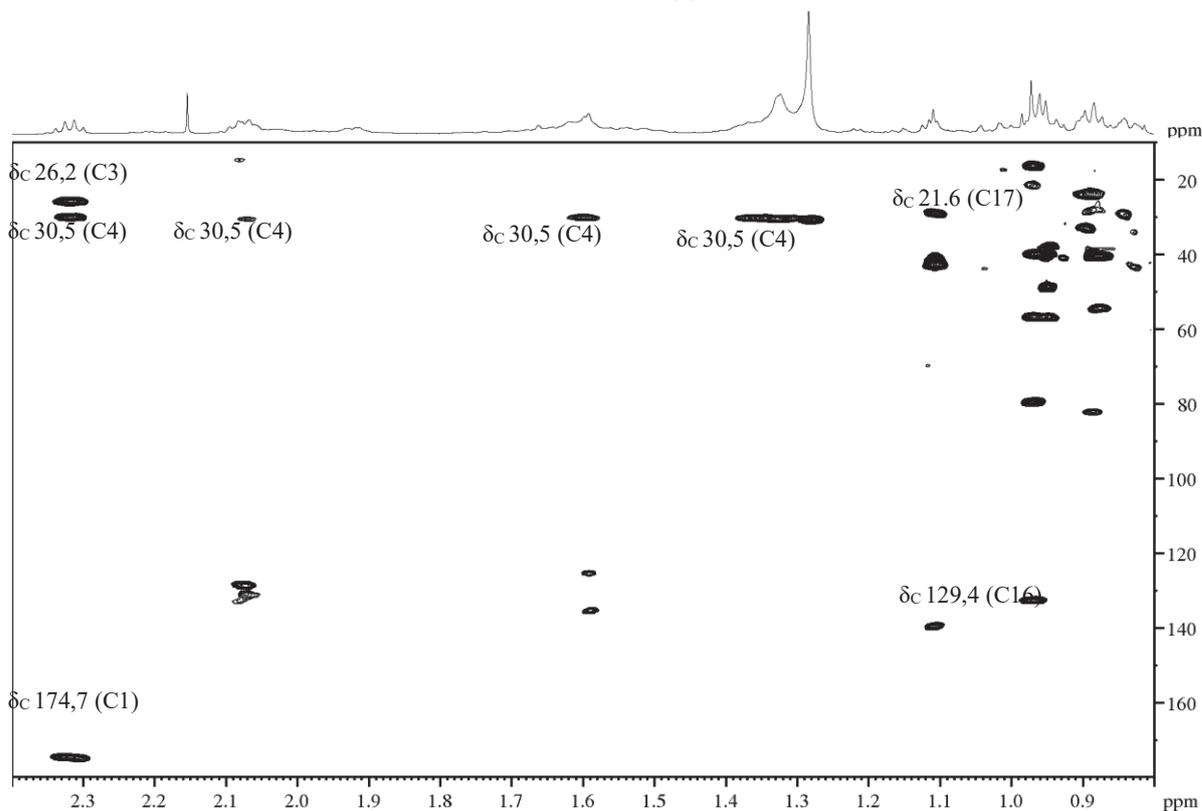
FONTE: A autora (2023).

FIGURA 52 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO DIRETA ^1H - ^{13}C (HSQC) DA AMOSTRA B38



FONTE: A autora (2023).

FIGURA 53 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO A LONGA DISTÂNCIA ^1H - ^{13}C (HMBC) DA AMOSTRA B38

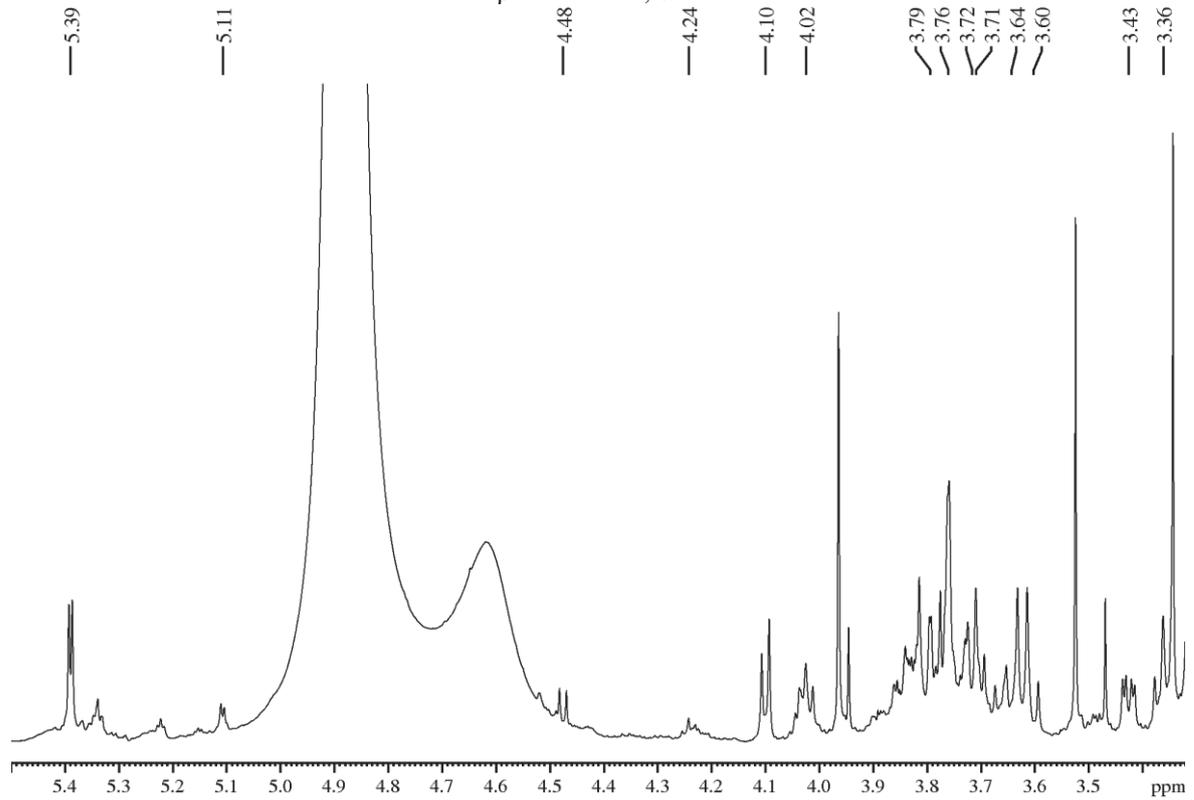


FONTE: A autora (2023).

5.2.5.2 Identificação estrutural da β -glucose (2), α -glucose (3) e sacarose (4)

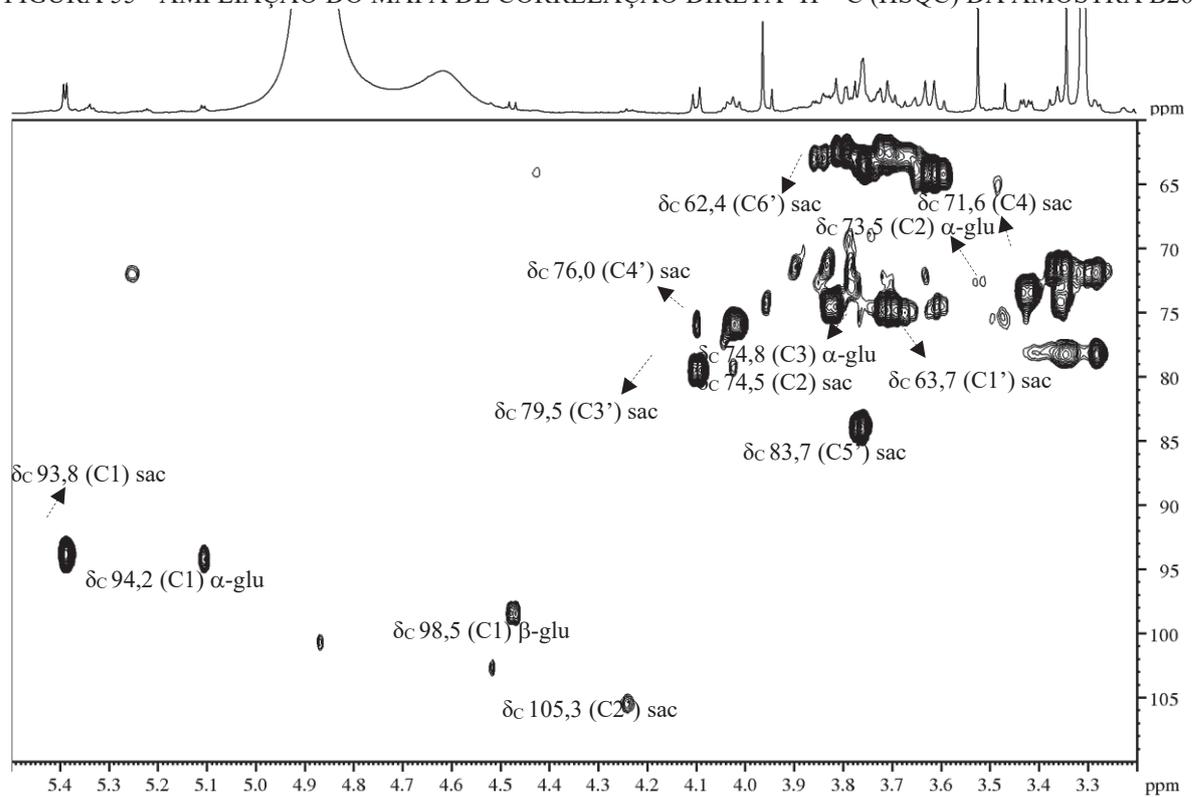
Na região entre δ_{H} 4,45-5,50 do espectro de RMN HR-MAS de ^1H foram observados três dupletos em δ_{H} 4,47 ($J= 7,7$), 5,11 ($J= 3,9$) e 5,39 ($J= 3,6$), relativos aos hidrogênios anoméricos (H1) dos açúcares β -glucose, α -glucose e sacarose. Estes sinais apresentaram correlação direta no mapa de correlação (^1H - ^{13}C) HSQC com os carbonos em δ_{C} 98,5, 94,2 e 93,8, respectivamente. Além disso, pôde-se observar que o sinal em 3,42 ppm correlaciona a longa distância com o carbono em δ_{C} 74,8 (C3), enquanto o sinal em 3,71 ppm correlaciona a longa distância com o carbono em δ_{C} 73,5 (C2), na α -glucose. Para a molécula da sacarose foi possível observar correlação a longa distância entre o H1 com C2 e C2', do H4 com o C2, dos H1' com C2' e C3', do H3' com C1' e C4', do H4' com os C3' e C6', e do H5' com o C4' (FIGURAS 54 a 56; TABELA 5) (TSUJIMOTO *et al.*, 2018).

FIGURA 54 - ESPECTRO DE RMN DE ^1H (δ_{H} 3,30-5,50) DA AMOSTRA B26 INDICANDO A PRESENÇA DOS CARBOIDRATOS β -GLUCOSE, α -GLUCOSE E SACAROSE



FONTE: A autora (2023).

FIGURA 55 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO DIRETA ^1H - ^{13}C (HSQC) DA AMOSTRA B26



FONTE: A autora (2023).

FIGURA 56 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO A LONGA DISTÂNCIA ^1H - ^{13}C (HMBC) DA AMOSTRA B26

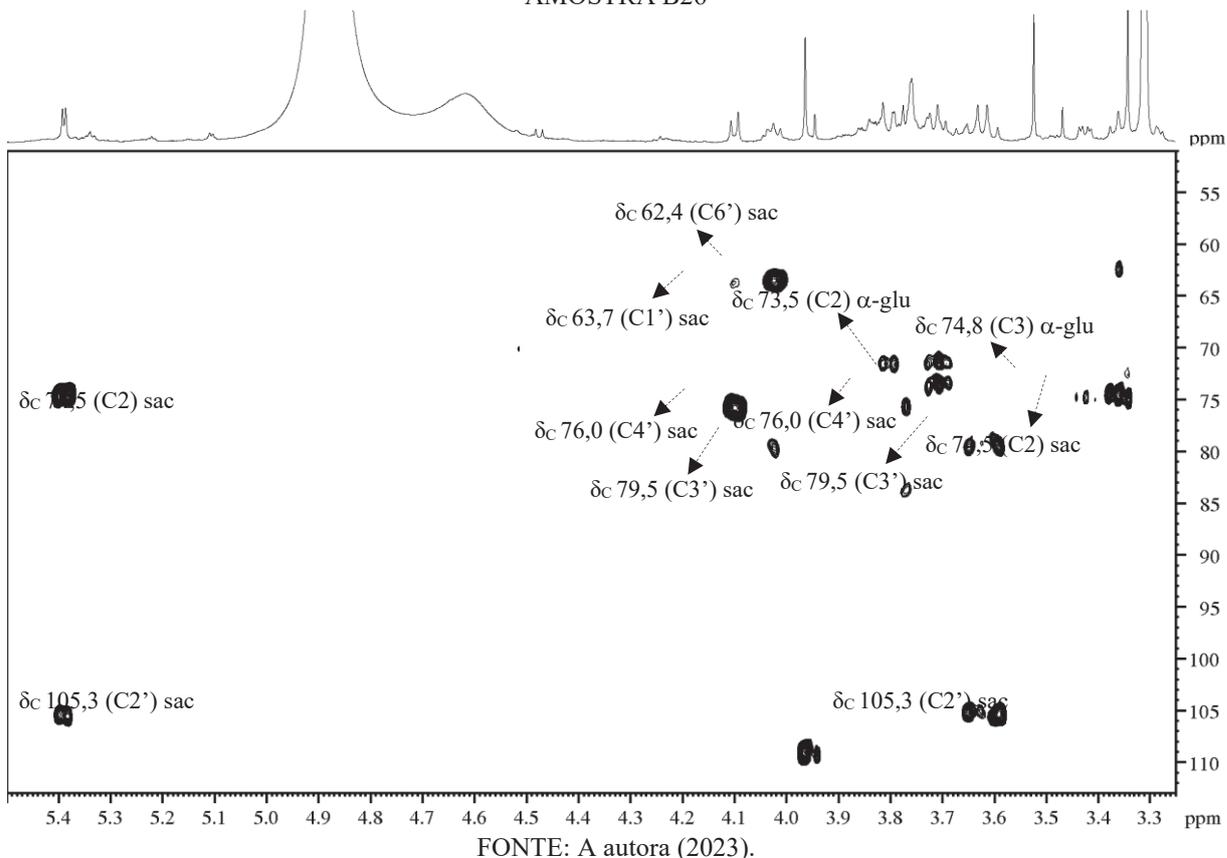
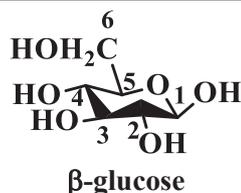


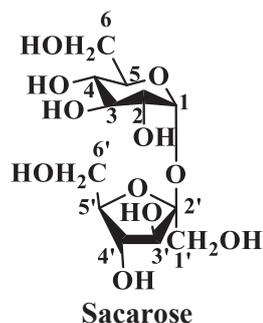
TABELA 3 - DADOS DE RMN HR-MAS DE ^1H E RMN 2D PARA A β -GLUCOSE, α -GLUCOSE E SACAROSE

Metabólito identificado			Literatura	
Posição	$\delta_{\text{H}}^{\text{a,b}}$	HSQC ^b	$\delta_{\text{H}}^{\text{c}}$	$\delta_{\text{C}}^{\text{c}}$
1	4,47 (<i>d</i> , $J=7,7$, 1H)	98,5	4,47 (<i>d</i> , $J=7,8$)	99,9

Metabólito identificado				Literatura	
Posição	$\delta_{\text{H}}^{\text{a,b}}$	HSQC ^b	HMBC ^b	$\delta_{\text{H}}^{\text{c}}$	$\delta_{\text{C}}^{\text{c}}$
1	5,11 (<i>d</i> , $J=3,9$, 1H)	94,2	-	5,09 (<i>d</i> , $J=3,7$)	95,7
2	3,42 (<i>dd</i> , $J=9,9/3,8$, 1H)	73,5	C3	3,34 (<i>dd</i> , $J=9,8/3,8$)	74,7
3	3,71 (<i>m</i> , 1H)	74,8	C2	-	-



Metabólito identificado			Literatura	
Posição	$\delta_H^{a,b}$	HSQC ^b	δ_H^c	δ_C^c



Metabólito identificado				Literatura	
Posição	$\delta_H^{a,b}$	HSQC ^b	HMBC ^b	δ_H^c	δ_C^c
1	5,39 (<i>d</i> , $J= 3,6$, 1H)	93,8	C2/C2'	5,37 (<i>d</i> , $J= 3,8$)	95,4
2	3,72 (<i>m</i> , 1H)	74,5	-	3,40 (<i>t</i> , $J= 9,8/3,8$)	74,9
3	-	-	-	3,68 (<i>t</i> , $J= 9,6$)	76,4
4	3,36 (<i>t</i> , $J= 9,3$, 1H)	71,6	C2	3,34 (<i>t</i> , $J= 9,4$)	73,0
5	-	-	-	-	76,1
6	-	-	-	3,70 (<i>dd</i> , $J= 7,9/4,0$)	63,9
1'	3,64 (<i>d</i> , $J= 12,4$, 1H)	63,7	C2'/C3'	3,62 (<i>d</i> , $J= 12,3$)	65,7
	3,60 (<i>d</i> , $J= 12,4$, 1H)			3,58 (<i>d</i> , $J= 12,3$)	
2'	4,24 (<i>m</i> , 1H)	105,3	-	-	107,1
3'	4,10 (<i>d</i> , $J= 8,3$, 1H)	79,5	C1'/C4'	4,08 (<i>d</i> , $J= 8,2$)	80,9
4'	4,02 (<i>t</i> , $J= 7,9$, 1H)	76,0	C3'/C6'	4,01 (<i>t</i> , $J= 7,7$)	77,4
5'	3,76 (<i>m</i> , 1H)	83,7	C4'	-	85,5
6'	3,79 (<i>dd</i> , $J= 11,1/2,1$, 1H)	62,4	-	3,79 (<i>dd</i> , $J= 12,0/2,7$)	65,7

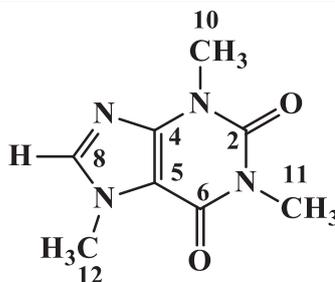
LEGENDA: ^AEXPERIMENTO REALIZADO A 400 MHZ PARA ¹H EM CD₃OD. ^BEXPERIMENTO REALIZADO A 600 MHZ PARA ¹H E 150 MHZ PARA ¹³C EM CD₃OD. ^{A,B}TMS COMO PADRÃO DE REFERÊNCIA. ^CTSUJIMOTO E COLABORADORES (2018) (800 MHZ PARA ¹H, CD₃OD/TMS).

5.2.5.3 Identificação estrutural da cafeína (5)

A análise dos perfis espectrais de RMN HR-MAS de ¹H das amostras possibilitaram observar três simpletos em δ_H 3,33, 3,51 e 3,95, relativos a três hidrogênios cada e, outro simpleto em δ_H 7,85 (1H). De acordo com o mapa de correlação direta (¹H-¹³C) HSQC esses sinais estão ligados aos carbonos em δ_C 28,4, 30,3, 34,1 e 144,3, respectivamente, referentes a três grupos N-metílicos (N-CH₃) e um hidrogênio aromático. A correta posição desses grupos foi estabelecida através do mapa de correlação a longa distância (¹H-¹³C) HMBC, no qual o sinal em 3,33 ppm correlacionou a longa distância com os carbonos em δ_C 153,4 (C2) e 157,3 (C6), o sinal em 3,51 ppm correlacionou a longa distância com os carbonos em δ_C 149,8 (C4) e 153,4 (C2), e o sinal em 3,95 ppm correlacionou a longa distância com os carbonos em δ_C 109,0

(C5) e 144,3 (C8); enquanto que o sinal em 7,85 ppm correlacionou a longa distância com os carbonos em δ_C 109,0 (C5) e 149,8 (C4) (FIGURAS 57 a 59; TABELA 6) (TAVARES e FERREIRA, 2006).

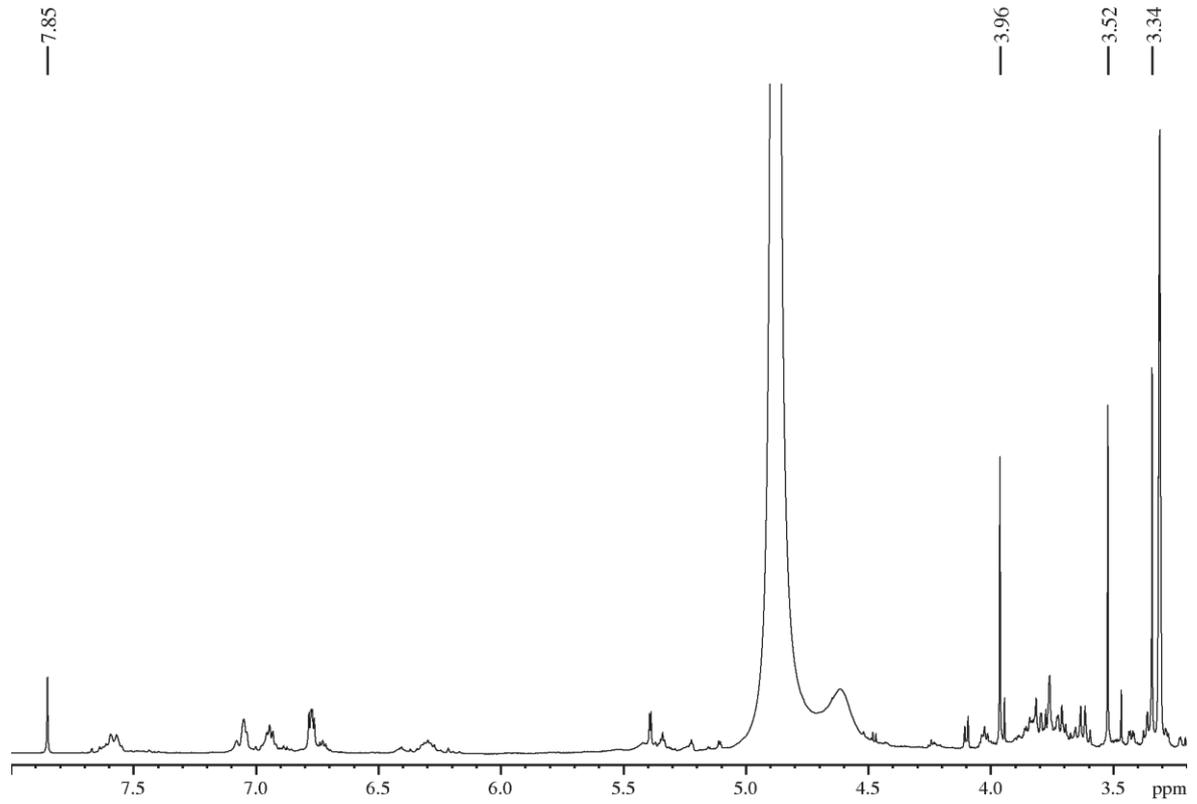
TABELA 4 - DADOS DE RMN HR-MAS DE ^1H E RMN 2D PARA A CAFEÍNA



Posição	Metabólito identificado			Literatura		
	$\delta_H^{a,b}$	HSQC ^b	HMBC ^b	δ_H^c	δ_C^c	HMBC ^c
1	-	-	-	-	-	-
2	-	153,4	-	-	153,8	-
3	-	-	-	-	-	-
4	-	149,8	-	-	149,8	-
5	-	109,0	-	-	108,0	-
6	-	157,3	-	-	158,0	-
7	-	-	-	-	-	-
8	7,85 (s, 1H)	144,3	C4/C5	7,89	144,8	C4/C5
9	-	-	-	-	-	-
10	3,51 (s, 3H)	30,3	C2/C4	3,48	30,9	C2/C4
11	3,33 (s, 3H)	28,4	C2/C6	3,30	28,9	C2/C6
12	3,95 (s, 3H)	34,1	C5/C8	3,91	34,5	C5/C8

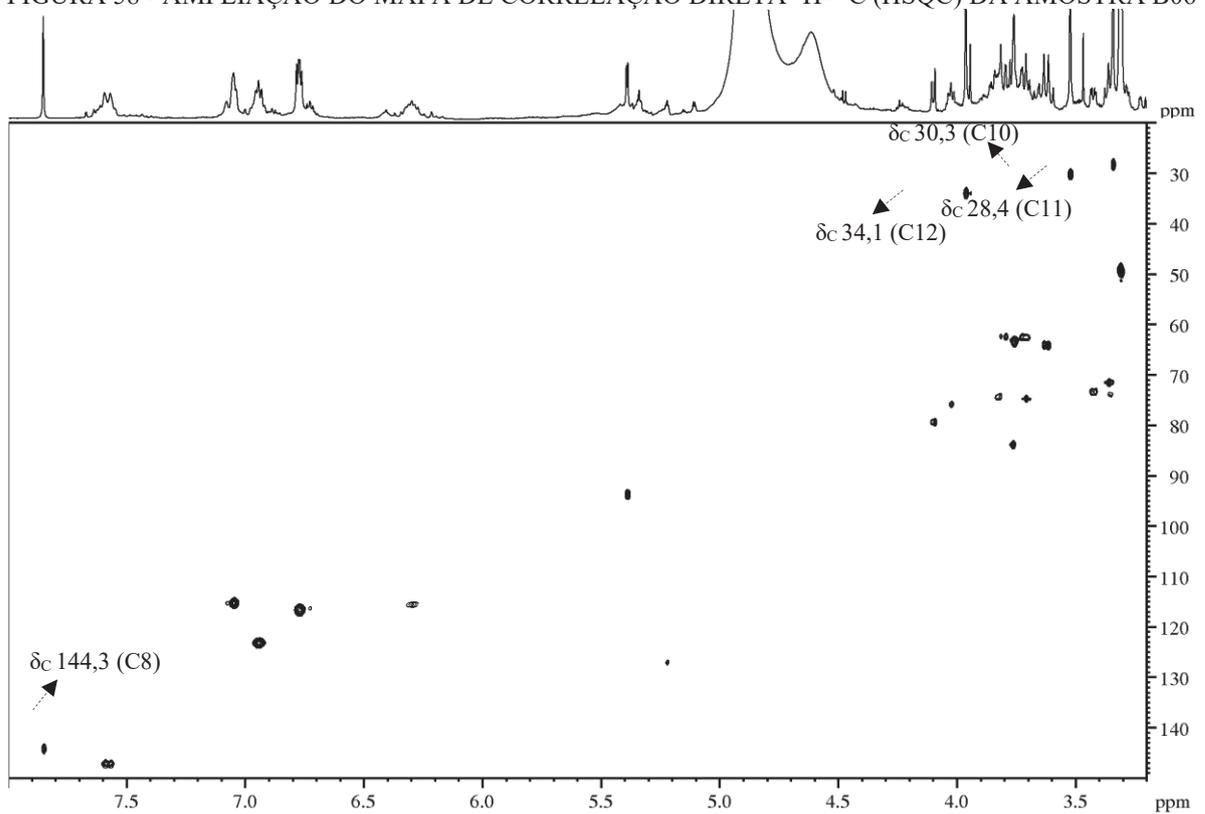
LEGENDA: ^AEXPERIMENTO REALIZADO A 400 MHZ PARA ^1H EM CD_3OD . ^BEXPERIMENTO REALIZADO A 600 MHZ PARA ^1H E 150 MHZ PARA ^{13}C EM CD_3OD . ^{A,B}TMS COMO PADRÃO DE REFERÊNCIA. ^CTAVARES E FERREIRA (2006) (400 MHZ PARA ^1H , $^{13}\text{C}\{^1\text{H}\}$ E GHMBC, $\text{DMSO-}D_6/\text{TMS}$).

FIGURA 57 - ESPECTRO DE RMN DE ^1H (δ_{H} 3,20-8,00) DA AMOSTRA B06 INDICANDO A PRESENÇA DE SINAIS CARACTERÍSTICOS PARA A CAFEÍNA



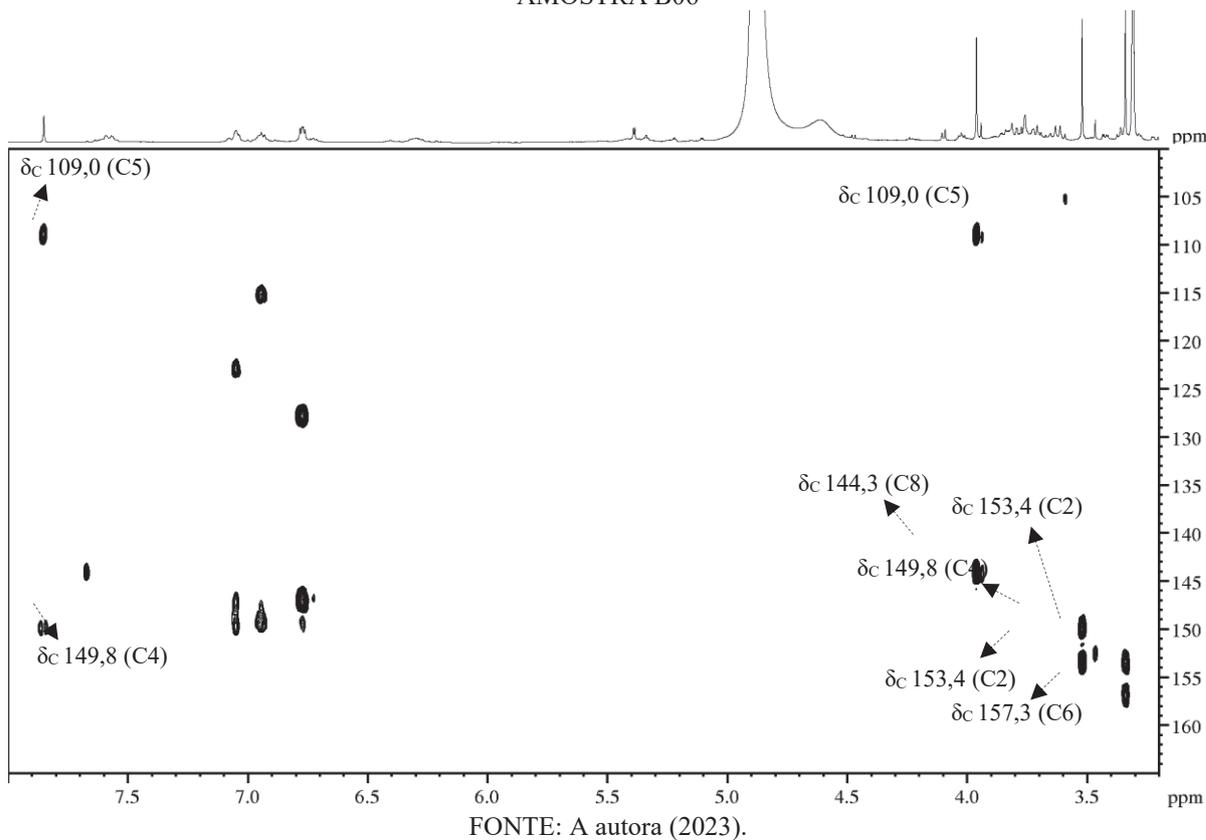
FONTE: A autora (2023).

FIGURA 58 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO DIRETA ^1H - ^{13}C (HSQC) DA AMOSTRA B06



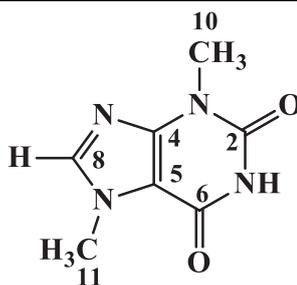
FONTE: A autora (2023).

FIGURA 59 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO A LONGA DISTÂNCIA ^1H - ^{13}C (HMBC) DA AMOSTRA B06



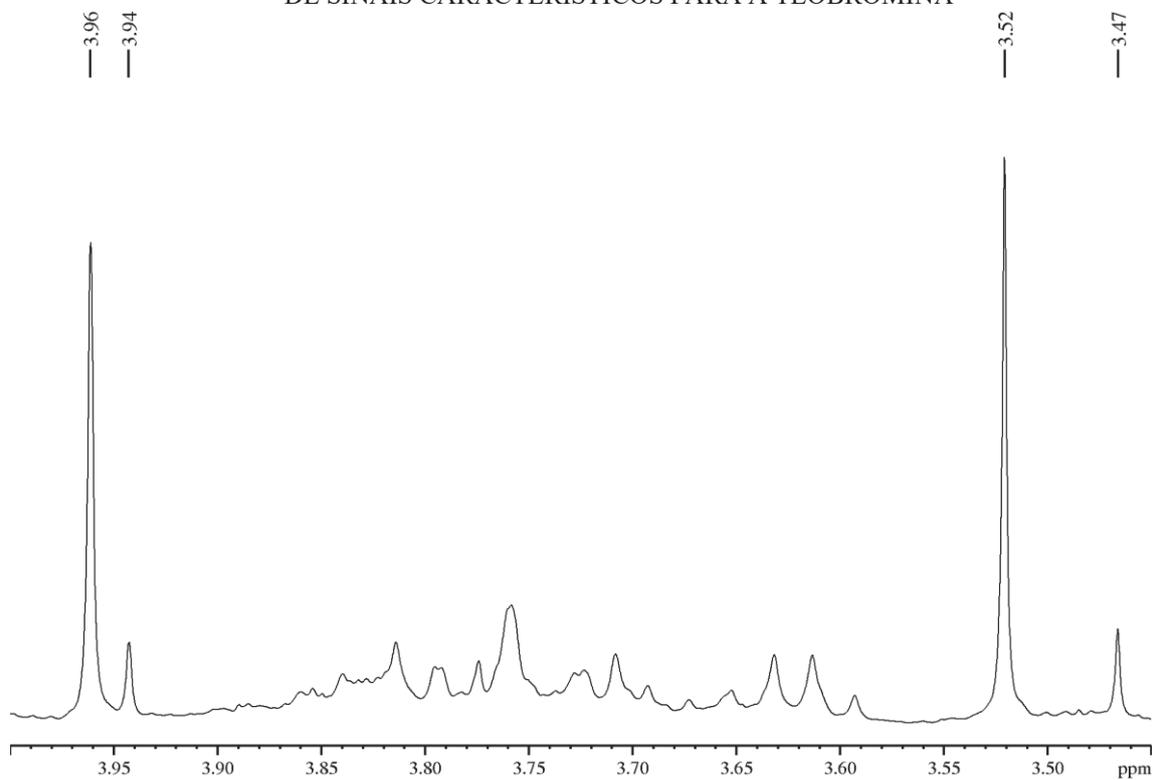
5.2.5.4 Identificação estrutural da teobromina (6)

A partir da análise dos perfis espectrais de RMN HR-MAS de ^1H pôde-se observar dois simpletos minoritários em δ_{H} 3,47 e 3,94, característicos dos hidrogênios dos grupos 3 e 7 N-metilícos do composto teobromina. De acordo com o mapa de correlação direta (^1H - ^{13}C) HSQC esses sinais estão ligados aos carbonos em δ_{C} 29,3 e 34,0, respectivamente. A correta posição desses grupos foi estabelecida através do mapa de correlação a longa distância (^1H - ^{13}C) HMBC, no qual o sinal em 3,46 ppm correlacionou a longa distância com o carbono em δ_{C} 152,5 (C2), e o sinal em 3,95 ppm correlacionou a longa distância com os carbonos em δ_{C} 109,0 (C5) e 144,3 (C8) (FIGURAS 60 a 62; TABELA 7) (FREITAS *et al.*, 2018).

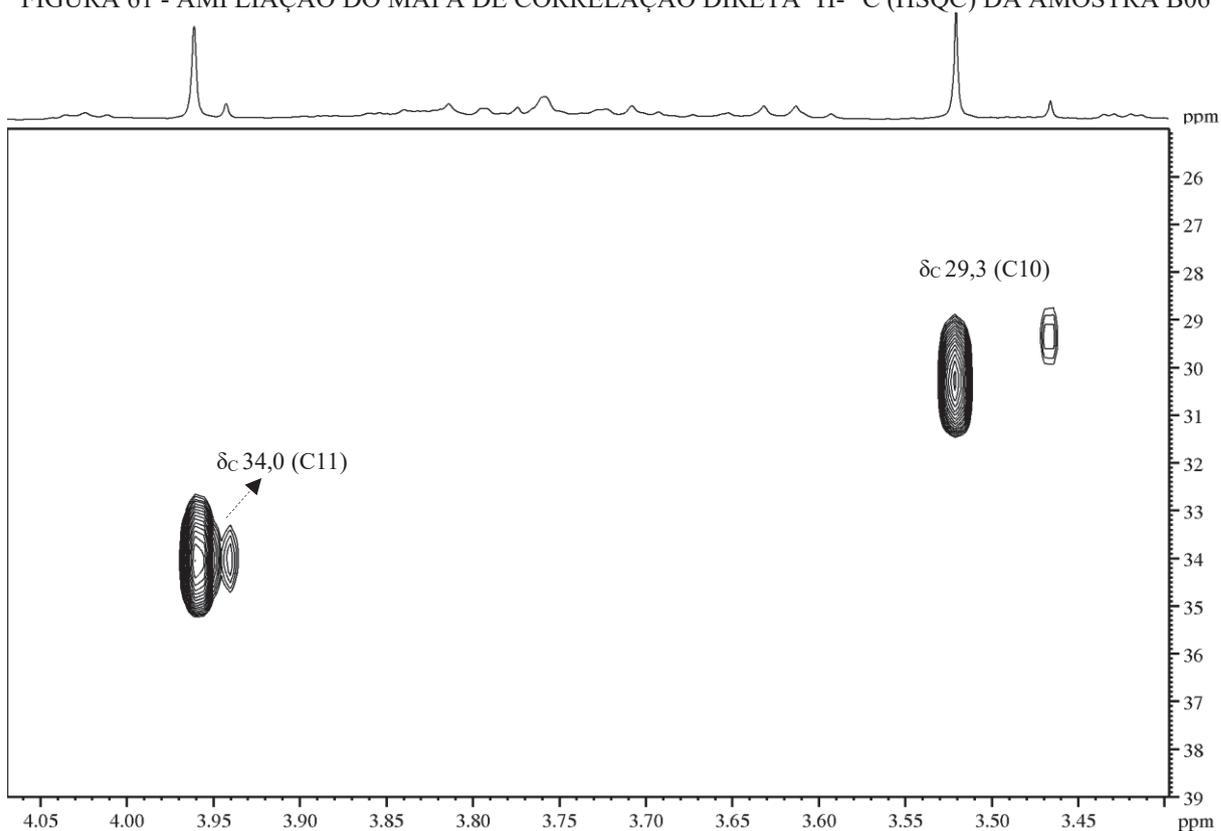
TABELA 5 - DADOS DE RMN HR-MAS DE ^1H E RMN 2D PARA A TEOBROMINA

Posição	Metabólito identificado			Literatura	
	$\delta_{\text{H}^{\text{a,b}}}$	HSQC ^b	HMBC ^b	$\delta_{\text{H}^{\text{c}}}$	$\delta_{\text{C}^{\text{c}}}$
1	-	-	-	-	-
2	-	152,5	-	-	-
3	-	-	-	-	-
4	-	-	-	-	-
5	-	109,0	-	-	-
6	-	-	-	-	-
7	-	-	-	-	-
8	-	144,3	-	8,00 (s)	144,4
9	-	-	-	-	-
10	3,47 (s, 3H)	29,3	C2	3,36 (s)	30,0
11	3,94 (s, 3H)	34,0	C5/C8	3,87 (s)	34,7

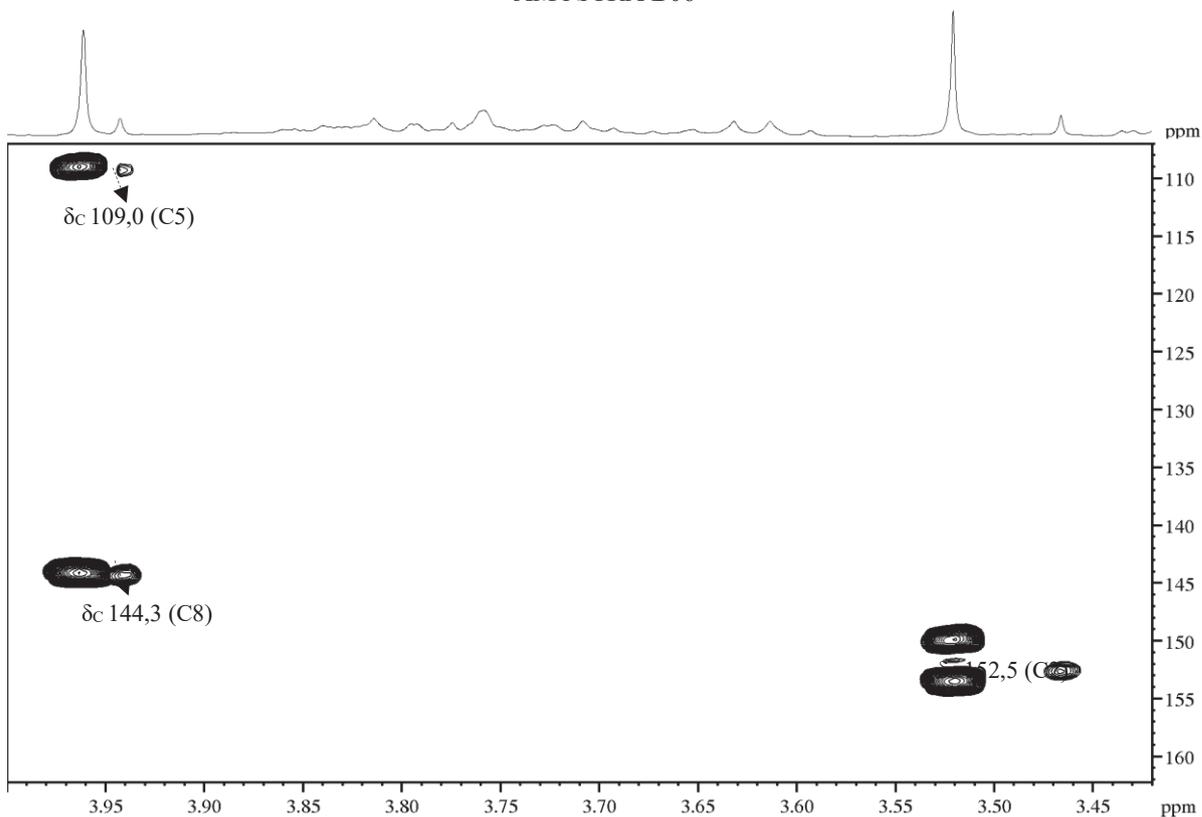
LEGENDA: ^AEXPERIMENTO REALIZADO A 400 MHZ PARA ^1H EM CD_3OD . ^BEXPERIMENTO REALIZADO A 600 MHZ PARA ^1H E 150 MHZ PARA ^{13}C EM CD_3OD . ^ABTMS COMO PADRÃO DE REFERÊNCIA. ^CFREITAS E COLABORADORES (2018) (300,13 MHZ PARA ^1H , $\text{D}_2\text{O}/\text{TMSP}$).

FIGURA 60 - ESPECTRO DE RMN DE ^1H (δ_{H} 3,40-4,00) DA AMOSTRA B06 INDICANDO A PRESENÇA DE SINAIS CARACTERÍSTICOS PARA A TEOBROMINA

FONTE: A autora (2023).

FIGURA 61 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO DIRETA ^1H - ^{13}C (HSQC) DA AMOSTRA B06

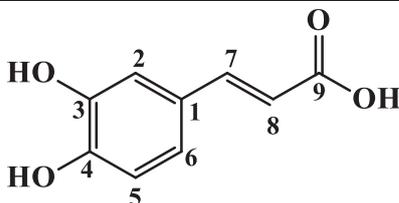
FONTE: A autora (2023).

FIGURA 62 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO A LONGA DISTÂNCIA ^1H - ^{13}C (HMBC) DA AMOSTRA B06

FONTE: A autora (2023).

5.2.5.5 Identificação estrutural do ácido cafeico (7)

A análise dos perfis espectrais de RMN HR-MAS de ^1H possibilitaram a visualização de dois dupletos característicos de uma ligação dupla do tipo *trans* (*E*) em δ_{H} 6,28 (*d*, $J= 15,9$, 1H) e 7,58 (*d*, $J= 15,9$, 1H). Esses sinais no mapa de correlação direta (HSQC) estão correlacionados aos carbonos em δ_{C} 115,7 (C8) e 147,3 (C7). No mapa de correlação a longa distância (HMBC), o sinal em 7,58 ppm correlacionou com os sinais em δ_{C} 115,3 (C2), 123,5 (C1) e 169,1 (C9). Além disso, observou-se sinais em δ_{H} 6,78 (*d*, $J= 7,7$, 1H), 6,94 (*m*, 1H) e 7,04 (*sl*, 1H), que correlacionaram com os carbonos em δ_{C} 116,6, 123,3 e 115,3, respectivamente. Analisando-se o mapa de correlação a longa distância (HMBC), verificou-se que o sinal em 6,78 ppm correlacionou com os carbonos em δ_{C} 128,0 (C1), 147,2 (C3) e 149,6 (C4); já o sinal em 6,94 ppm correlacionou com os carbonos em δ_{C} 115,3 (C2), 147,2 (C3) e 149,6 (C4); enquanto o sinal em 7,04 ppm correlacionou com os carbonos em 123,3 (C6), 147,2 (C3) e 149,6 (C4). De acordo com esses dados o composto foi identificado como sendo o ácido cafeico (FIGURAS 63 a 65; TABELA 8) (JEONG *et al.*, 2011).

TABELA 6 - DADOS DE RMN HR-MAS DE ^1H E RMN 2D PARA O ÁCIDO CAFEICO

Posição	Metabólito identificado			Literatura	
	$\delta_{\text{H}}^{\text{a,b}}$	HSQC ^b	HMBC ^b	$\delta_{\text{H}}^{\text{c}}$	$\delta_{\text{C}}^{\text{c}}$
1	-	123,5	-	-	128,3
2	7,04 (<i>sl</i> , 1H)	115,3	C3/C4/C6	7,07 (<i>d</i> , $J= 2,0$, 1H)	115,7
3	-	147,2	-	-	147,2
4	-	149,6	-	-	149,8
5	6,78 (<i>d</i> , $J= 7,7$, 1H)	116,6	C1/C3/C4	6,81 (<i>d</i> , $J= 8,2$, 1H)	117,0
6	6,94 (<i>m</i> , 1H)	123,3	C2/C3/C4	6,95 (<i>dd</i> , $J= 8,0/2,0$, 1H)	123,4
7	7,58 (<i>d</i> , $J= 15,9$, 1H)	147,3	C1/C2/C9	7,55 (<i>d</i> , $J= 15,9$, 1H)	147,6
8	6,28 (<i>d</i> , $J= 15,9$, 1H)	115,7	-	6,24 (<i>d</i> , $J= 15,9$, 1H)	1165,0
9	-	169,1	-	-	171,6

LEGENDA: ^AEXPERIMENTO REALIZADO A 400 MHZ PARA ^1H EM CD_3OD . ^BEXPERIMENTO REALIZADO A 600 MHZ PARA ^1H E 150 MHZ PARA ^{13}C EM CD_3OD . ^{A,B}TMS COMO PADRÃO DE REFERÊNCIA. ^CJEONG E COLABORADORES (2011) (500 MHZ PARA ^1H E 125 MHZ PARA ^{13}C , $\text{CD}_3\text{OD}/\text{TMS}$).

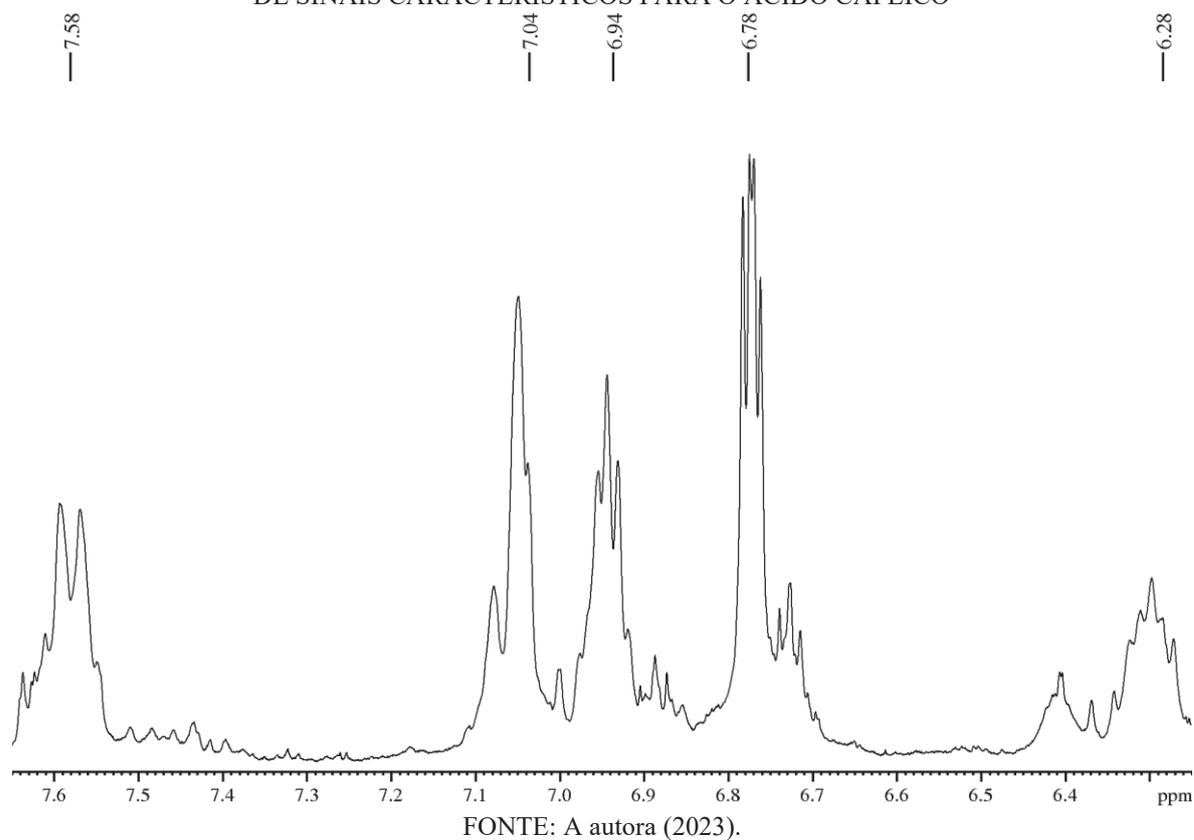
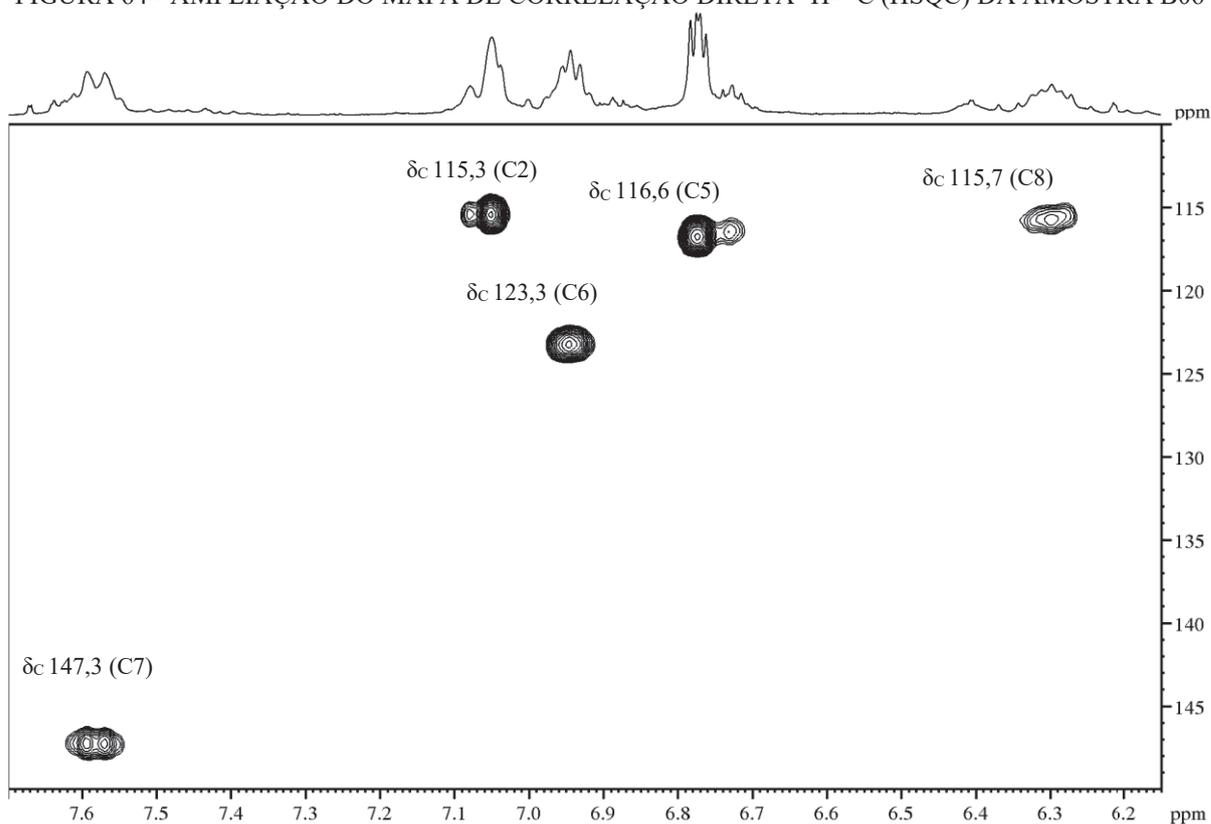
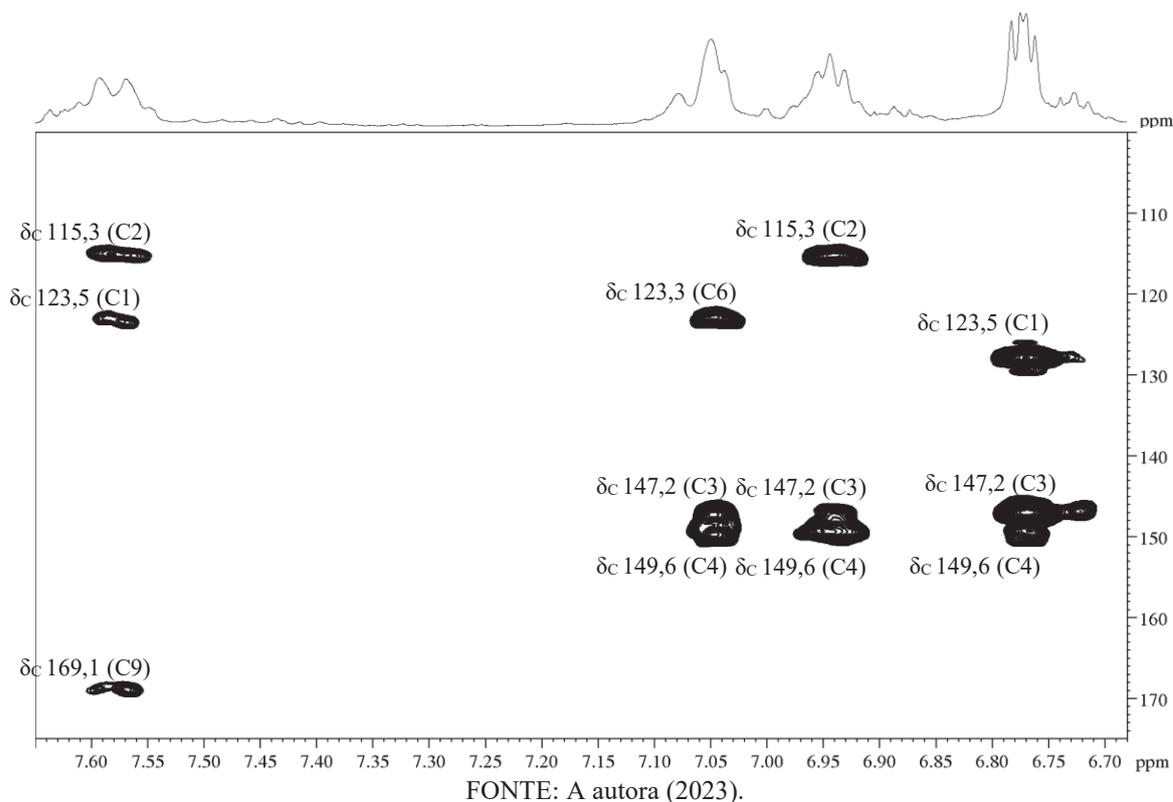
FIGURA 63 - ESPECTRO DE RMN DE ^1H (Δ_{H} 6,25-7,65) DA AMOSTRA B06 INDICANDO A PRESENÇA DE SINAIS CARACTERÍSTICOS PARA O ÁCIDO CAFEICOFIGURA 64 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO DIRETA ^1H - ^{13}C (HSQC) DA AMOSTRA B06

FIGURA 65 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO A LONGA DISTÂNCIA ^1H - ^{13}C (HMBC) DA AMOSTRA B06



5.2.5.6 Identificação estrutural do aminoácido GABA (8)

A análise do perfil espectral possibilitou a visualização e identificação do ácido γ -aminobutírico (GABA), isso devido aos sinais em δ_H 1,90 (*m*, 2H) e 2,31 (*t*, $J=7,3$, 2H). Os valores de deslocamento químico dos carbonos com correlação direta aos sinais dos hidrogênios do metabólito acima apresentados foram observados no mapa de correlação HSQC (^1H - ^{13}C) e os de correlação a longa distância através do mapa de correlação HMBC (^1H - ^{13}C) (FIGURAS 66 e 67; TABELA 9) (OLIVEIRA *et al.*, 2014).

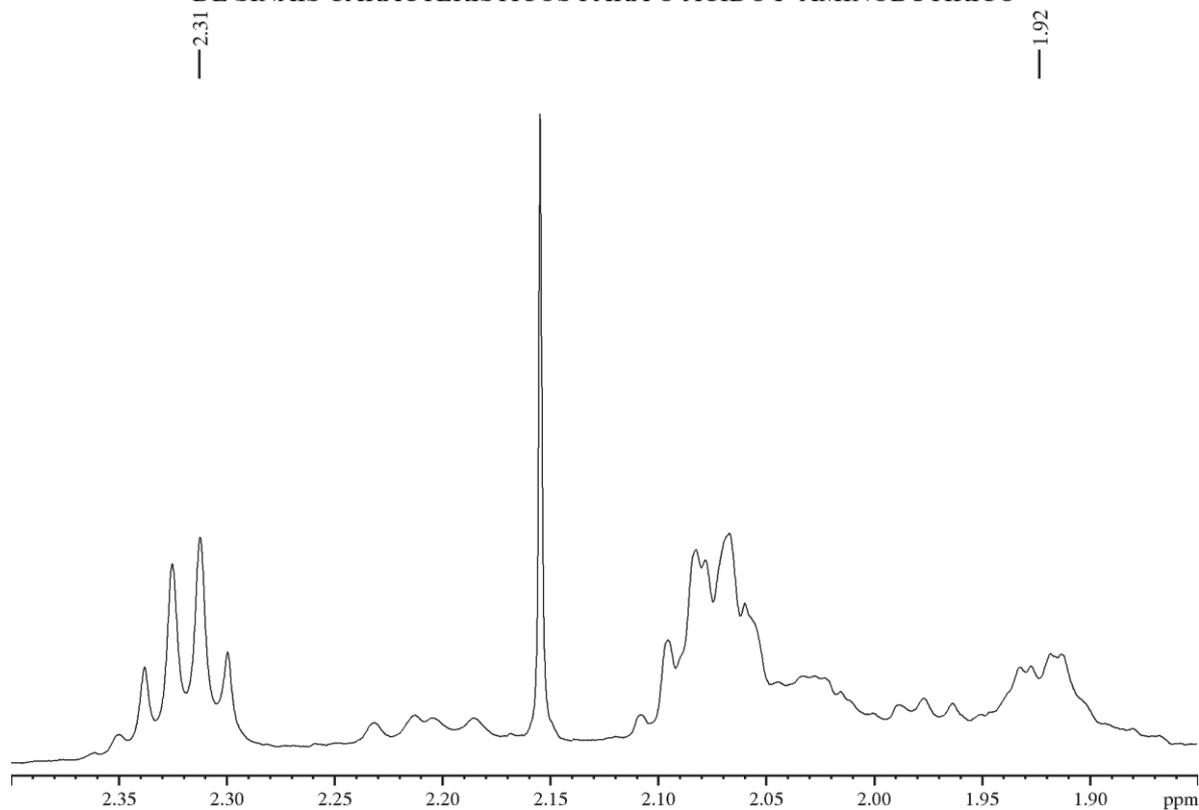
TABELA 7 - DADOS DE RMN HR-MAS DE ^1H E RMN 2D PARA O ÁCIDO γ -AMINOBTIRICO

NCCCC(=O)O

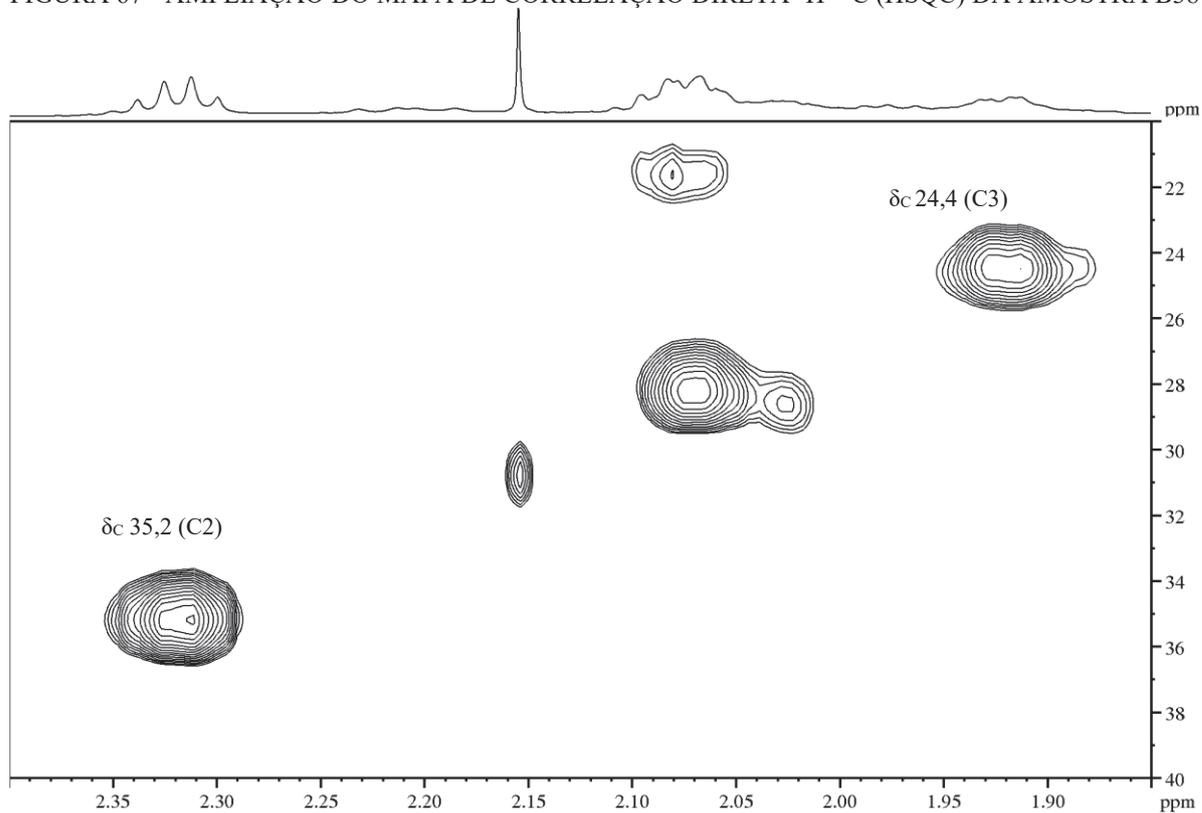
H₂N 4 3 2 1 OH

Posição	Metabólito identificado		Literatura	
	$\delta_H^{a,b}$	HSQC ^b	δ_H^c	δ_C^c
1	-	-	-	-
2	2,31 (<i>t</i> , $J=7,6$, 2H)	35,2	2,32 (<i>t</i> , $J=7,24$)	37,0
3	1,92 (<i>m</i> , 2H)	24,4	1,91	-
4	-	-	3,02 (<i>t</i> , $J=7,46$)	42,0

LEGENDA: ^AEXPERIMENTO REALIZADO A 400 MHZ PARA ^1H EM CD_3OD . ^BEXPERIMENTO REALIZADO A 600 MHZ PARA ^1H E 150 MHZ PARA ^{13}C EM CD_3OD . ^{A,B}TMS COMO PADRÃO DE REFERÊNCIA. ^COLIVEIRA E COLABORADORES (2014) (500 MHZ PARA ^1H E 125 MHZ PARA ^{13}C , $\text{D}_2\text{O}/\text{TMS}$).

FIGURA 66 - ESPECTRO DE RMN DE ^1H (δ_{H} 1,85-2,40) DA AMOSTRA B38 INDICANDO A PRESENÇA DE SINAIS CARACTERÍSTICOS PARA O ÁCIDO Γ -AMINO BUTÍRICO

FONTE: A autora (2023).

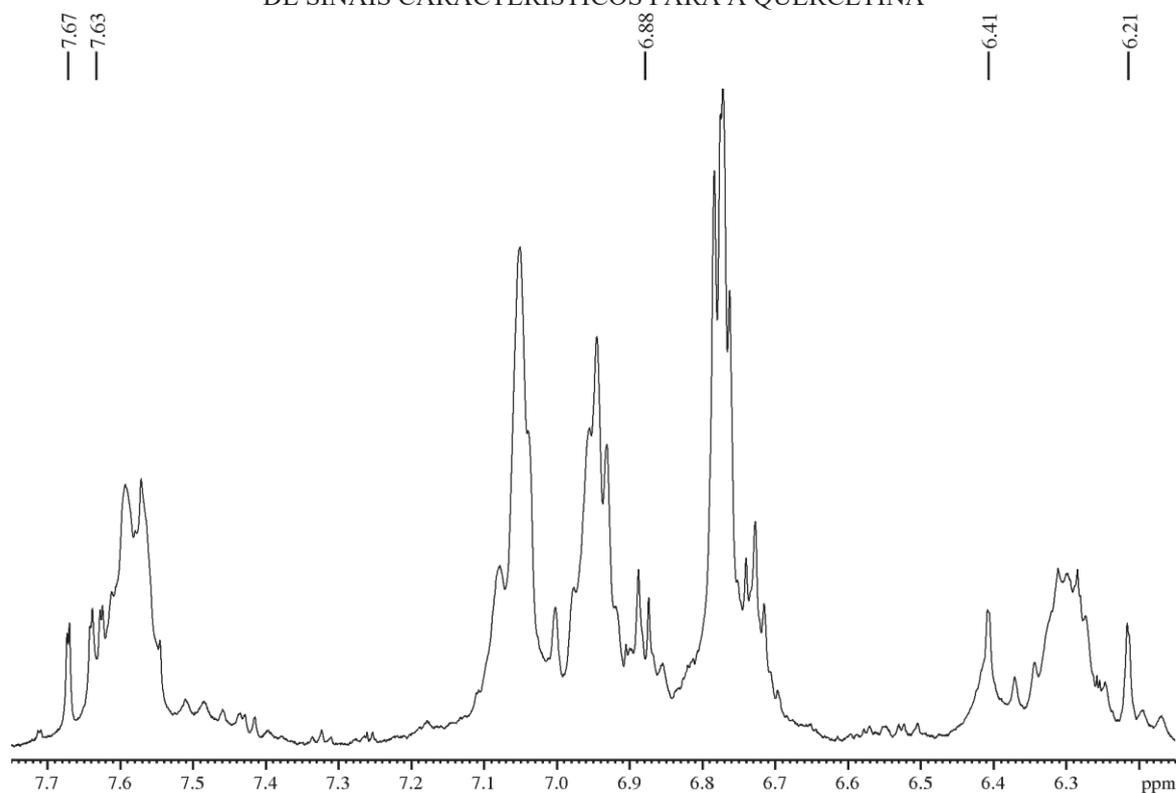
FIGURA 67 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO DIRETA ^1H - ^{13}C (HSQC) DA AMOSTRA B38

FONTE: A autora (2023).

5.2.5.7 Identificação estrutural da quercetina (9)

Nos perfis espectras foi possível observar dois pares de dupletos em δ_H 6,21 ($d, J= 2,0, 1H$) e 6,41 ($d, J= 2,0, 1H$), característicos dos hidrogênios aromáticos de acoplamento *meta* do flavonoide quercetina, os quais apresentaram correlação direta com os carbonos em δ_C 100,0 (C6) e 95,1 (C8), respectivamente. Foram observados a presença de dois dupletos em 6,88 ($d, J= 8,4, 1H$) e 7,67 ($d, J= 2,0, 1H$), e um duplo duplete em δ 7,63 ($dd, J= 8,4/2,0, 1H$), os quais apresentaram correlação direta com os carbonos em δ_C 116,1 (C5'), 117,9 (C2') e 123,8 (C6'). Analisando-se o mapa de correlação a longa distância (HMBC), verificou-se que o sinal em 6,21 ppm correlacionou com os carbonos em δ_C 95,1 (C8) e 105,9 (C10), o sinal em 6,41 ppm correlacionou com os carbonos em δ_C 105,9 (C10) e 158,5 (C9), o sinal em 6,88 ppm correlacionou com os carbonos em δ_C 122,9 (C1') e 146,4 (C2), o sinal em 7,63 ppm correlacionou com os carbonos em δ_C 115,4 (C3') e 117,9 (C2'), enquanto que o sinal em 7,67 ppm correlacionou com os carbonos em δ_C 122,9 (C1'), 144,1 (C3') e 149,9 (C4') (FIGURAS 68 a 70; TABELA 10) (Sa *et al.*, 2017).

FIGURA 68 - ESPECTRO DE RMN DE 1H (δ_H 6,25-7,65) DA AMOSTRA B48 INDICANDO A PRESENÇA DE SINAIS CARACTERÍSTICOS PARA A QUERCETINA



FONTE: A autora (2023).

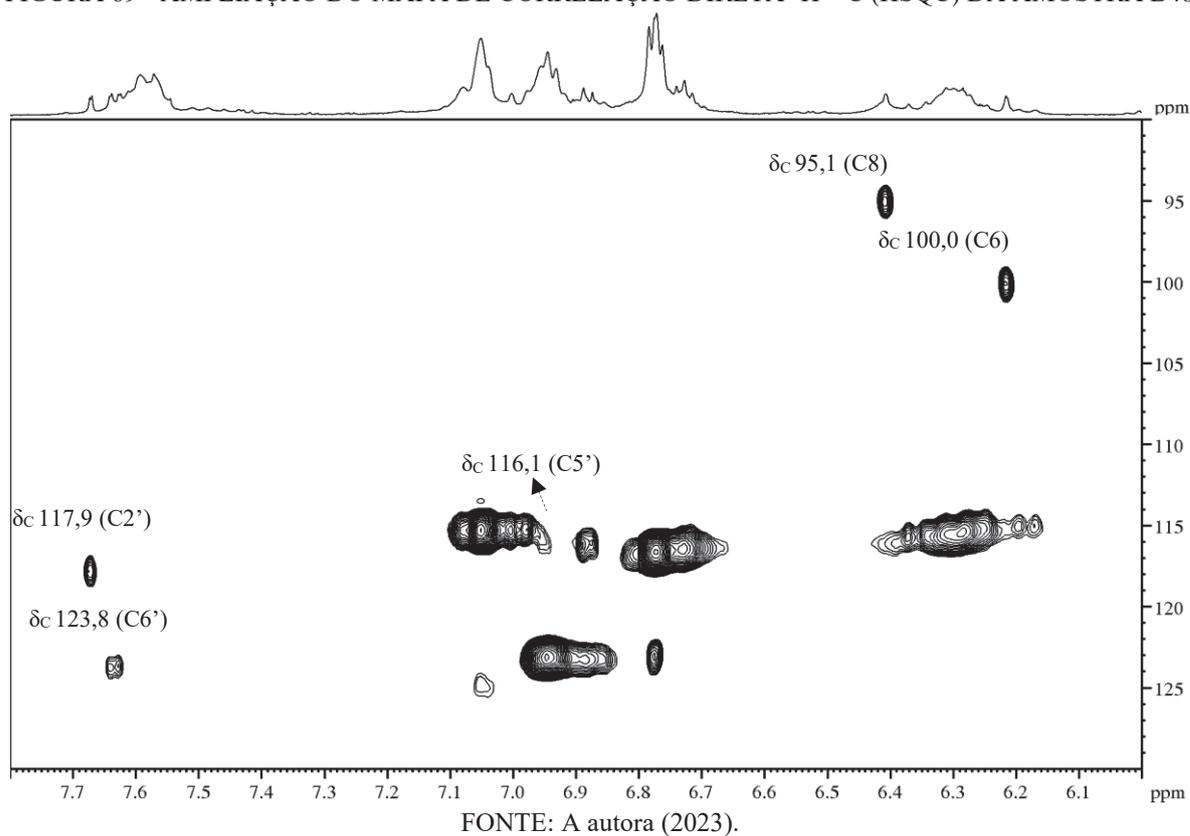
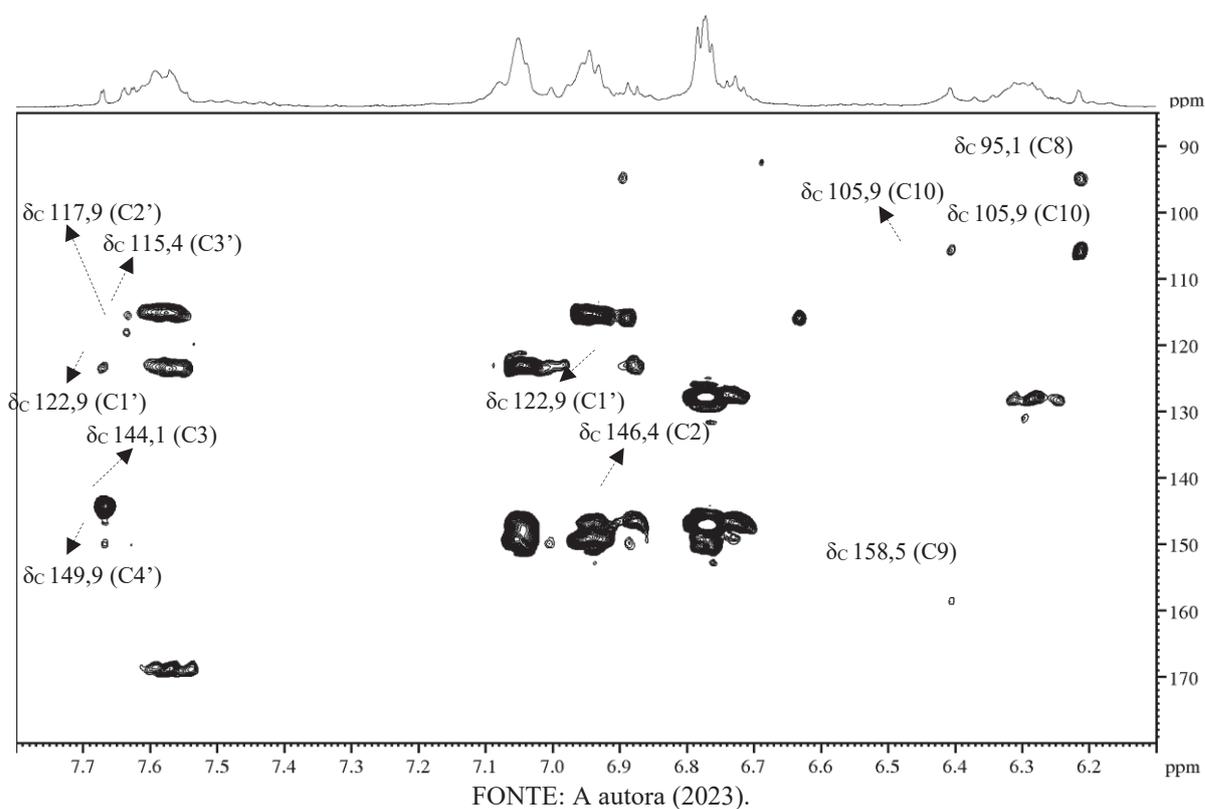
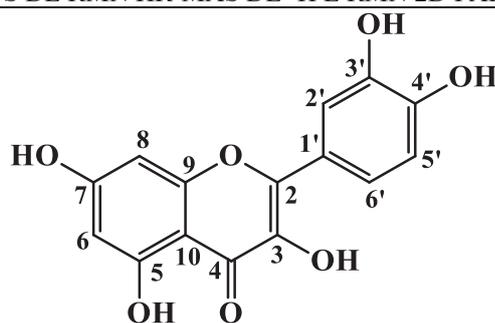
FIGURA 69 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO DIRETA ^1H - ^{13}C (HSQC) DA AMOSTRA B48FIGURA 70 - AMPLIAÇÃO DO MAPA DE CORRELAÇÃO A LONGA DISTÂNCIA ^1H - ^{13}C (HMBC) DA AMOSTRA B48

TABELA 8 - DADOS DE RMN HR-MAS DE ¹H E RMN 2D PARA A QUERCETINA

Posição	Metabólito identificado			Literatura	
	$\delta_H^{a,b}$	HSQC ^b	HMBC ^b	δ_H^c	δ_C^c
1	-	-	-	-	-
2	-	146,4	-	-	147,1
3	-	144,1	-	-	137,8
4	-	-	-	-	175,3
5	-	-	-	-	163,4
6	6,21 (<i>d</i> , <i>J</i> = 2,0, 1H)	100,0	C8/C10	6,18 (<i>d</i> , <i>J</i> = 1,9, 1H)	97,8
7	-	-	-	-	166,1
8	6,41 (<i>d</i> , <i>J</i> = 2,0, 1H)	95,1	C9/C10	6,39 (<i>d</i> , <i>J</i> = 2,1, 1H)	93,1
9	-	158,5	-	-	156,6
10	-	105,9	-	-	102,7
1'	-	122,9	-	-	122,2
2'	7,67 (<i>d</i> , <i>J</i> = 2,0, 1H)	117,9	C1'/C4'	7,73 (<i>d</i> , <i>J</i> = 2,1, 1H)	114,6
3'	-	115,4	-	-	114,0
4'	-	149,9	-	-	147,6
5'	6,88 (<i>d</i> , <i>J</i> = 8,4, 1H)	116,1	C2'/C1'	6,88 (<i>d</i> , <i>J</i> = 8,4, 1H)	114,7
6'	7,63 (<i>dd</i> , <i>J</i> = 8,4/2,0, 1H)	123,8	C2'/C3'	7,61 (<i>dd</i> , <i>J</i> = 8,4/2,1, 1H)	120,4

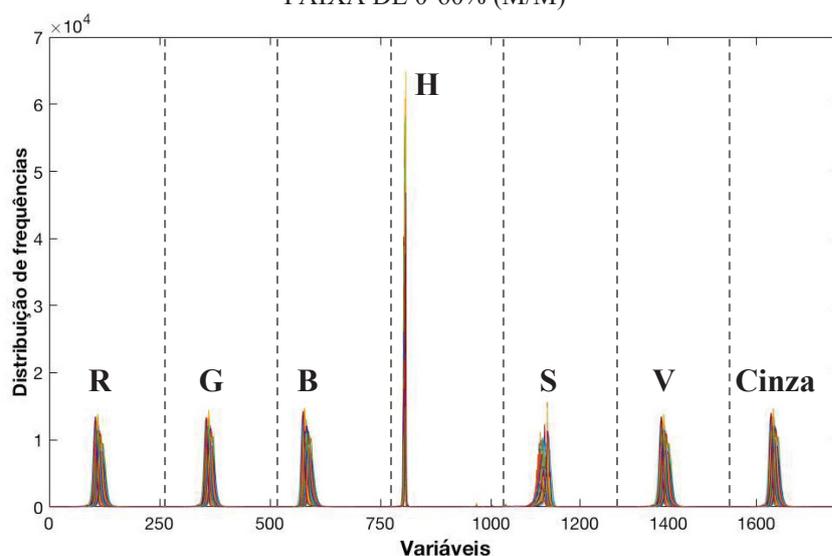
LEGENDA: ^aEXPERIMENTO REALIZADO A 400 MHZ PARA ¹H EM CD₃OD. ^bEXPERIMENTO REALIZADO A 600 MHZ PARA ¹H E 150 MHZ PARA ¹³C EM CD₃OD. ^{a,b}TMS COMO PADRÃO DE REFERÊNCIA. ^cSA E COLABORADORES (2017) (500 MHZ PARA ¹H E 125 MHZ PARA ¹³C, CD₃OD/TMS).

5.3 QUIMIOMETRIA APLICADA À ANÁLISE DE IMAGENS DIGITAIS

5.3.1 Quantificação de talos em amostras de erva-mate comerciais via regressão PLS

Com o intuito de desenvolver uma metodologia para a quantificação da porcentagem de talos presente em amostras de erva-mate, o algoritmo *Kennard-Stone* foi aplicado ao banco de dados contendo os espectros NIR das amostras de erva-mate (A01-A82) para a seleção de 5 amostras que fossem representativas da variabilidade dos dados. Em seguida, essas amostras foram utilizadas para o preparo das misturas com talo na faixa de 0-60% (m/m) (conforme o procedimento descrito na seção 3.7.3 desse documento) e as imagens digitais foram coletadas. Na FIGURA 71, são apresentados os histogramas de cor das 168 amostras (misturas + triplicatas) utilizadas na construção dos modelos.

FIGURA 71 - HISTOGRAMAS DE COR DE AMOSTRAS DE ERVA-MATE ADICIONADAS DE TALO NA FAIXA DE 0-60% (M/M)



FONTE: A autora (2023).

O banco de dados contendo os histogramas combinados dos canais de cor RGB, HSV e escala de cinza contendo as 168 amostras, foi subdividido com o auxílio do algoritmo *Kennard-Stone* em dois conjuntos, o de calibração contendo 134 amostras e o de validação externa, contendo 34 amostras. Vale ressaltar que o algoritmo de seleção foi aplicado de forma a manter todas as réplicas das amostras dentro do mesmo conjunto (calibração ou validação), evitando assim, o sobreajuste dos modelos construídos.

Modelos sobreajustados (*overfitted*) apresentam alto desempenho para as amostras do conjunto de calibração, porém não são capazes de gerar previsões acuradas para novos conjuntos amostras. Uma das estratégias de detecção de sobreajuste nos modelos multivariados é a comparação entre os valores de RMSE, visto que um aumento significativo de RMSEP pode indicar que o modelo apresenta baixa capacidade de generalização (FABER & RAJKÓ, 2007).

A fim de avaliar a influência de cada espaço de cor, a construção dos modelos PLS foi realizada empregando-se tanto espaços únicos quanto o histogramas combinados dos canais RGB, HSV e escala de cinza, sendo que os pré-processamentos avaliados variaram de acordo com as características da região utilizada na modelagem.

Devido às diferenças de escala apresentadas pelo canal H (matiz), ao utilizar a região HSV ou o histograma compilado, os métodos de escalamento (*e.g.* autoescalamento e escalamento de Pareto) foram avaliados a fim de evitar que o algoritmo de PLS considerasse um peso maior para o canal H (matiz) e garantir uma avaliação mais precisa das contribuições das variáveis no modelo matemático. Por outro lado, os canais RGB e escala de cinza foram avaliados com a aplicação dos dados centrados na média.

Na TABELA 11, são apresentadas as principais informações sobre os modelos (*i.e.* região do histograma utilizada e pré-processamento) e as suas respectivas métricas de avaliação de qualidade (*i.e.* número de VL, R^2_{CV} , R^2_P , RMSECV e RMSEP). Os valores de R^2 para o conjunto de predição variaram entre 0,25 (escala de cinzas) até 0,83 (região HSV) enquanto os valores de RMSEP variaram entre 6,41 (região HSV) e 15,25% (escala de cinzas).

A partir desses resultados, foi possível concluir que o espaço de cor HSV (destacado na TABELA 11) apresentou as menores variações entre os valores de R^2 e RMSE obtidos com as amostras de calibração e de predição. Isso indica que o modelo apresenta um maior potencial de resultar em predições acuradas quando aplicado a novos conjuntos de amostras.

TABELA 9 - Parâmetros de desempenho dos modelos de regressão pls para a quantificação de talos em amostras de erva-mate comercial a partir de imagens digitais

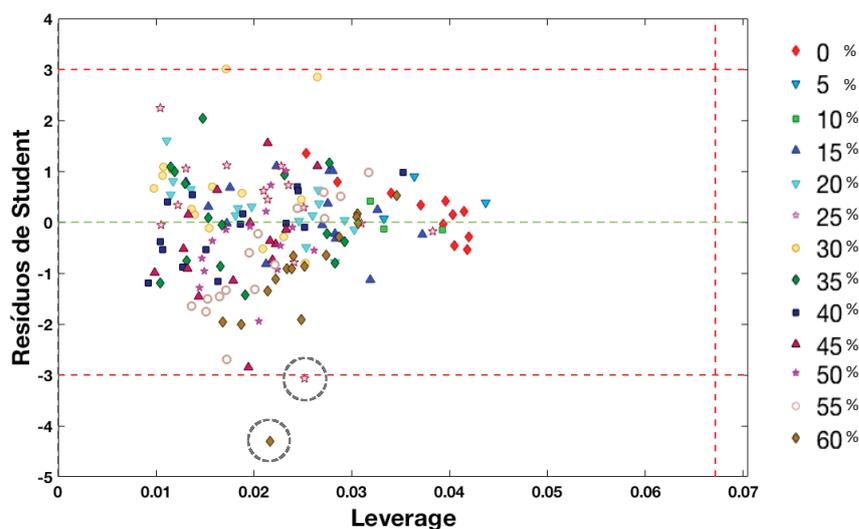
Região	Pré-processamento	VL	R^2_{CV}	R^2_P	RMSECV	RMSEP
R+G+B	Centrados na média	3	0,85	0,75	6,70	7,92
H+S+V	Autoescalamento	5	0,91	0,77	5,14	7,28
	Escalamento de Pareto	3	0,88	0,83	5,75	6,41
Escala de cinzas	Centrados na média	2	0,57	0,25	11,23	15,28
RGB+HSV	Autoescalamento	5	0,91	0,77	5,24	7,10
	Escalamento de Pareto	3	0,87	0,80	6,09	7,01
Histograma concatenado	Autoescalamento	5	0,90	0,77	5,31	7,12
	Escalamento de Pareto	3	0,87	0,80	6,19	7,19

FONTE: A autora (2023).

O modelo construído com a região de escala de cinzas não apresentou ajuste linear satisfatório e resultou em erros de predição elevados. Além disso, a combinação entre o espaço HSV e os demais, não resultou em RMSEP reduzidos, indicando que as informações existentes nesses canais não contribuem significativamente para a obtenção de modelos mais acurados.

A detecção de amostras com comportamento anômalo foi efetuada durante a etapa de otimização dos modelos de calibração através da avaliação dos valores de resíduos de *Student* e *leverage* (FIGURA 72). Duas amostras (circuladas no gráfico) apresentaram valores de resíduo fora dos limites, o que indica que os valores preditos pelo modelo não se aproximaram do valor declarado como real. O comportamento anômalo pode ter sido causado por erros no preparo da replicata ou interferências na imagem coletada, sendo assim, essas amostras foram removidas e as métricas de desempenho foram reavaliadas.

FIGURA 72 - RESÍDUOS DE *STUDENT VS. LEVERAGE*: MODELO PLS PARA QUANTIFICAÇÃO DE TALO EM ERVA-MATE



FONTE: A autora (2023).

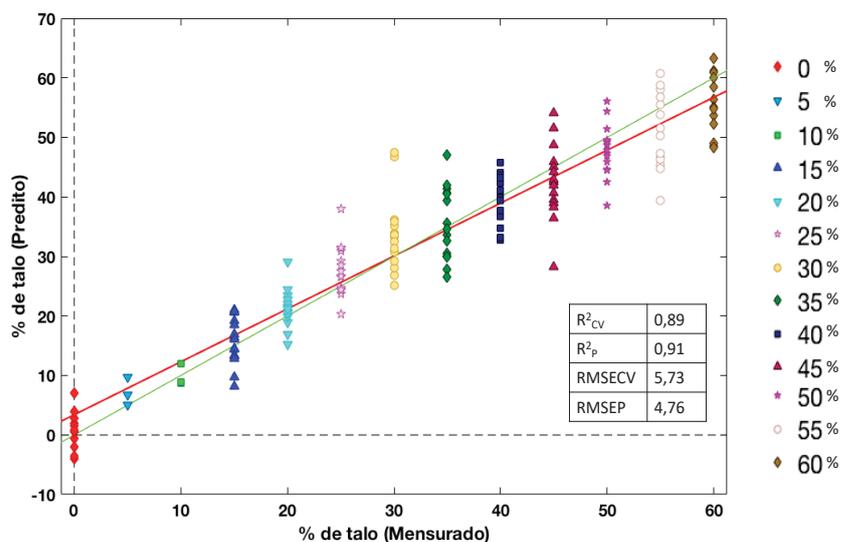
Como pode ser observado na FIGURA 73, a remoção das amostras contendo altos valores de resíduo resultou na diminuição do erro de predição e em aumento no valor do coeficiente de determinação tanto para o conjunto de calibração quanto para as amostras de predição. O modelo apresentou linearidade na faixa estudada, 0-60% (m/m), o que permitiria a aplicação para as amostras de erva-mate comerciais que segundo a classificação PN podem apresentar entre 30-50% de talos em sua composição.

Esses resultados indicam que a metodologia de quantificação de talos em amostras de erva-mate através da análise de imagens digitais, sugerida nesse estudo, apresenta potencial para ser aplicada no controle de qualidade do produto, visto que se trata de uma estratégia rápida, de baixo custo, e que apresentou erros de predição abaixo de 10% (TABELA 12).

No gráfico de VIP (FIGURA 74), podem ser identificadas as variáveis que mais contribuíram para a classificação das amostras, sendo possível constatar que, apesar de todos os canais (H, S e V) apresentarem valores de significância acima do *threshold*, o canal H é o que apresenta os maiores valores de scores. Conforme mencionado na seção 2.7 desse documento, o canal H (matiz) descreve a cor básica ou pureza de tonalidade da imagem.

No gráfico de scores da FIGURA 75 podem ser observados dois agrupamentos em relação aos valores de PC1. As amostras contendo menores quantidades de talo (<30%) geram imagens que apresentam uma coloração verde mais uniforme e conseqüentemente valores mais altos de matiz. Por outro lado, amostras com maior quantidade de talo ($\geq 30\%$) apresentam menor pureza de tonalidade, o que resulta em menores valores no canal H.

FIGURA 73 - VALORES MENSURADOS VS PREDITOS: MODELO PLS PARA QUANTIFICAÇÃO DE TALO EM ERVA-MATE



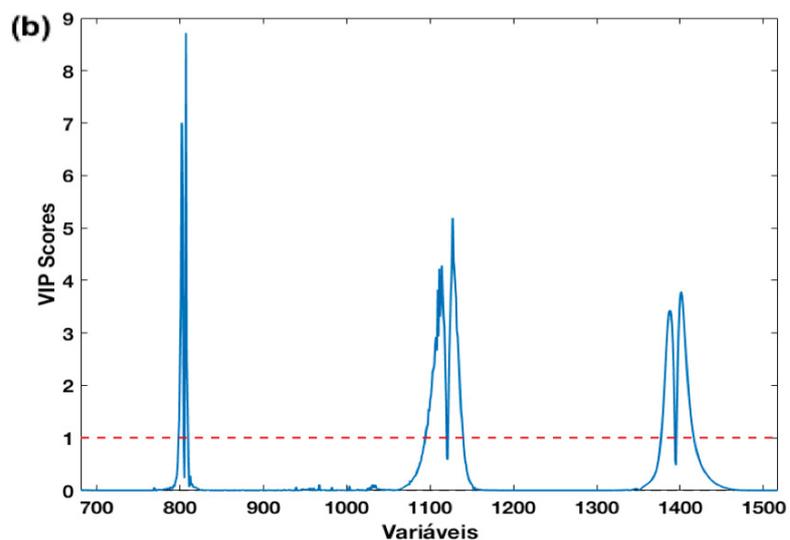
FONTE: A autora (2023).

TABELA 10 - Média e desvio-padrão dos valores preditos de % de talo nas amostras do conjunto de validação

Amostra	% de talo adicionado	% de talo predito (média, n=3)	Desvio padrão
A39	15	44,10	4,94
A39	40	42,35	2,62
A39	60	57,46	2,23
A40	50	44,10	4,94
A40	60	46,25	9,66
A46	20	22,34	1,81
A46	40	39,19	3,47
A52	35	37,37	3,99
A52	45	42,52	3,73
A52	55	47,46	2,74
A72	55	56,43	2,47

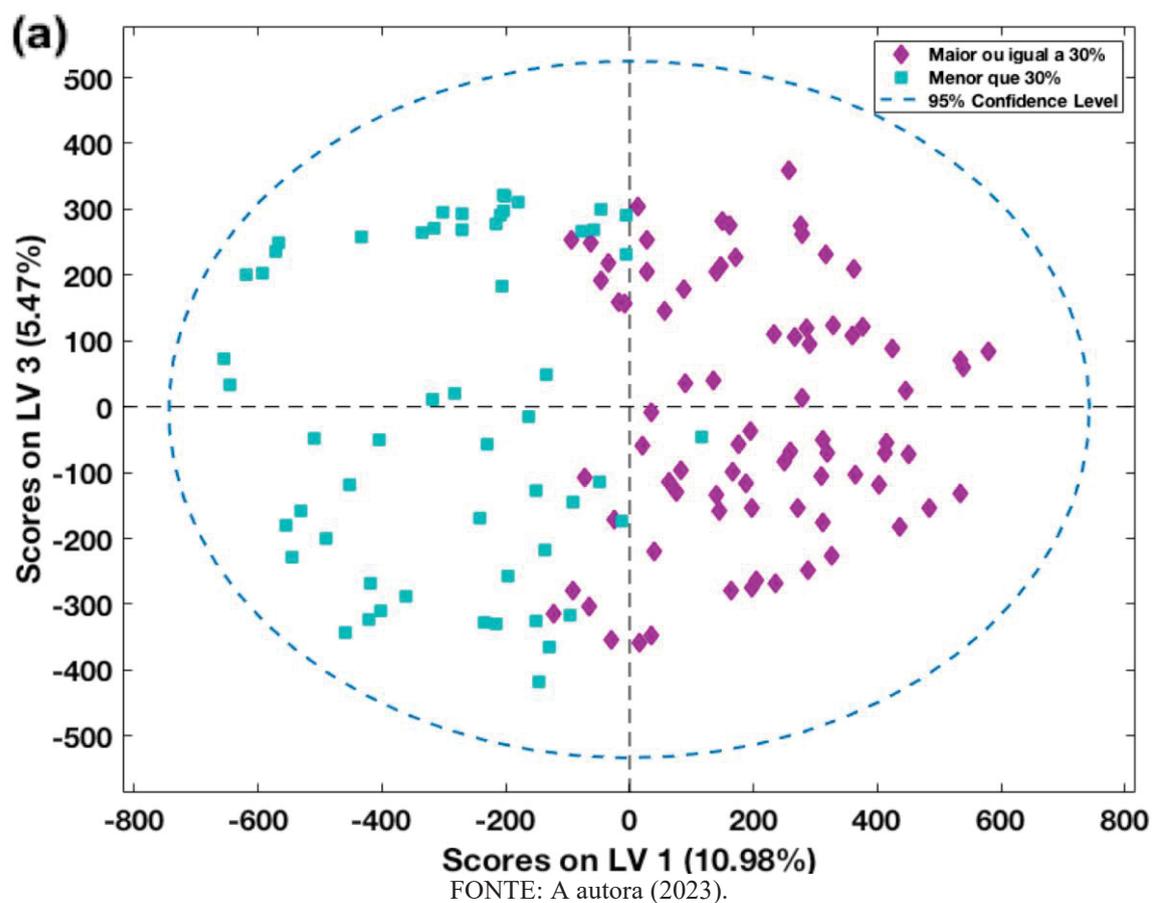
FONTE: A autora (2023).

FIGURA 74 - VIP SCORES DO MODELO PLS CONSTRUÍDO COM 3 VL'S A PARTIR DOS DADOS DE IMAGENS DIGITAIS PARA QUANTIFICAÇÃO DE TALO EM ERVA-MATE



FONTE: A autora (2023).

FIGURA 75 - GRÁFICOS DE SCORES (PC1 VS PC3) OBTIDO A PARTIR DOS DADOS DE IMAGENS DIGITAIS DAS AMOSTRAS DE ERVA-MATE IDENTIFICADAS DE ACORDO COM A FAIXA DE CONCENTRAÇÃO DE TALO



Apesar das vantagens de utilização dos métodos baseados em imagens digitais, uma consulta bibliográfica mostra que essa estratégia ainda é pouco aplicada em estudos envolvendo matrizes vegetais, sendo mais comum o uso da técnica de imagens hiperespectrais, que é capaz de fornecer informações espaciais e químicas das amostras (SUN et al., 2022; MAO et al., 2022; LIU, 2022; DJOKAM et al., 2017). Entretanto, as câmeras hiperespectrais apresentam desvantagens, como custo elevado e necessidade de maior tempo de treinamento dos analistas, o que justifica a busca por metodologias alternativas.

6 CONCLUSÕES

A aplicação da PCA nos dados de NIR e MIR possibilitou a identificação de tendências de agrupamento baseados o tipo de produto, confirmando a possibilidade de identificação de amostras com adição de sacarose. Além disso, pode ser verificado o agrupamento das amostras baseado no tipo de embalagem, indicando que ocorrem alterações significativas na composição química de matrizes alimentícias devido ao aumento de permeação causado pelo uso de embalagens de papel ou transparentes.

A PCA aplicada aos dados espectrais NIR em dois períodos distintos, corroborou a importância da definição da janela de tempo entre a abertura da embalagem e as análises para que não ocorram diferenças significativas no perfil espectral das amostras. Esse resultado também indica que a técnica poderia ser aplicada em estudos de tempo de prateleira, principalmente relacionados a condições que gerem aumento de umidade.

A partir da análise exploratória dos dados de infravermelho, não foi possível verificar agrupamentos relacionados com a origem geográfica das amostras pois, apesar de alguns estudos terem demonstrado o potencial das técnicas multivariadas no agrupamento de amostras de erva-mate de acordo com os países de origem, a similaridade das condições climáticas, solo e práticas agrícolas dentro do mesmo país faz com que a discriminação seja dificultada.

A associação entre a espectroscopia NIR e o método de reconhecimento de padrões supervisionado PLS-DA possibilitou a construção de um modelo de classificação para amostras comerciais de erva-mate contendo sacarose, que pode ser uma alternativa de metodologia rápida e sem preparo de amostra para o controle de qualidade desse tipo de produto. Entretanto, a fim de aumentar a robustez do modelo e evitar tendências causadas pela presença de classes desbalanceadas, recomenda-se a adição de um maior número de amostras adoçadas.

A PCA aplicada aos dados de RMN ^1H possibilitou a discriminação das amostras produzidas na região de São Mateus do Sul, tendo sido identificadas as regiões que apresentam deslocamentos químicos que podem ser atribuídos à presença de cafeína, compostos fenólicos e saponinas, como os principais fatores de diferenciação com relação às amostras provenientes de outras origens geográficas. A partir desse resultado, foi possível construir um modelo de classificação PLS-DA que resultou em uma taxa de classificações corretas acima de 78%.

A análise do perfil espectral de RMN HR-MAS de ^1H e de RMN em solução uni e bidimensionais, bem como comparações com dados da literatura, permitiu identificar 9 marcadores químicos que influenciam significativamente na discriminação das amostras de

acordo com a origem geográfica, sendo estes, o ácido linolênico, a β -glucose, a α -glucose, a sacarose, a cafeína, a teobromina, o ácido cafeico, o aminoácido GABA e a quercetina.

Devido ao fato de alimentos de origem vegetal apresentarem grande variabilidade, recomenda-se a adição de um maior número de amostras aos modelos construídos. Além disso, para fins de gerar evidências para a obtenção do selo de denominação de origem, recomenda-se a elucidação dos compostos responsáveis pela discriminação a partir de RMN bidimensional.

Os dados de imagens digitais obtidos com o scanner de mesa, possibilitaram a construção de um modelo de calibração multivariada para a quantificação de talo nas amostras de erva-mate. Além do baixo custo e de não requerer preparo da amostra, a metodologia apresentou alta frequência analítica, possibilitando a coleta das imagens de 24 amostras de erva-mate de maneira simultânea. Uma alternativa para aumentar ainda mais a frequência analítica, é a otimização do aparato com placas de menor diâmetro.

De modo geral, as metodologias descritas nesse documento podem ser acrescidas de robustez a partir do incremento do número de amostras avaliadas, o que possibilitaria melhor captura da variabilidade intrínseca de produtos de origem vegetal. Entretanto, vale ressaltar que atualmente não existem metodologias de análise direta da erva-mate comercial para a determinação da quantidade de talos, presença de sacarose ou para a discriminação da origem geográfica à nível de microrregiões ou cidades.

Sendo assim, os bons resultados obtidos para as estratégias rápidas e com mínimo preparo de amostra desenvolvidas durante esse estudo, apresentam potencial para serem aplicadas e contribuir significativamente com o controle de qualidade de amostras de erva-mate. Além disso, essas metodologias podem auxiliar na obtenção de informações relevantes sobre os compostos que caracterizam as amostras da região de indicação de procedência, o que pode vir a ser utilizado para a potencial obtenção do selo de denominação de origem.

REFERÊNCIAS

- ABBAS, O. et al. Phenolic compound explorer: a mid-infrared spectroscopy database. **Vibrational Spectroscopy**, v. 92, p. 111-118, 2017.
- ALAM, T. M.; JENKINS, J. E. HR-MAS NMR spectroscopy in material science. **Advanced aspects of spectroscopy**, v. 10, p. 279-306, 2012.
- ALAMSJAH, M. A.; HIRAO, S.; ISHIBASHI, F.; FUJITA, Y. Isolation and structure determination of algicidal compounds from *Ulva fasciata*. **Bioscience, Biotechnology, and Biochemistry**, v. 69, n. 11, p. 2186–2192, 2005.
- ALMEIDA, A. G. et al. Portable NIR Spectroscopy-Chemometric Identification of Chemically Differentiated Yerba Mate (*Ilex paraguariensis*) Clones. **Food Analytical Methods**, v. 16, n. 3, p. 469-477, 2023.
- ALMEIDA, F. M. N. **Espectroscopia de Infravermelho Próximo com Transformada de Fourier (FT-NIR) na Caracterização de Farinhas para Alimentação Pueril**. 84 f. Dissertação (Mestrado em Engenharia Biológica) - Instituto Superior Técnico, Universidade Técnica De Lisboa, Lisboa, 2009.
- ANNADURAI, S. **Fundamentals of digital image processing**. Pearson Education India, 2007.
- ANTONIADIS, P. How to Convert an RGB Image to a Grayscale. 19 de jun. de 2023. Disponível em: <https://www.baeldung.com/cs/convert-rgb-to-grayscale>. Acesso em: 2 de jul de 2023.
- ASHTON, R. H. “Nothing good ever came from New Jersey”: Expectations and the sensory perception of wines. **Journal of Wine Economics**, v. 9, n. 3, p. 304-319, 2014.
- AUGUSTIJN, D.; DE GROOT, Huub JM; ALIA, A. HR-MAS NMR applications in plant metabolomics. **Molecules**, v. 26, n. 4, p. 931-949, 2021.
- AZEVEDO, R. S. A.; TEIXEIRA, B. S.; SAUTHIER, M. C. DA S.; et al. Multivariate analysis of the composition of bioactive in tea of the species *Camellia sinensis*. **Food Chemistry**, v. 273, n. April 2018, p. 39–44, 2019.
- BAJOUB, A. et al. Assessing the varietal origin of extra-virgin olive oil using liquid chromatography fingerprints of phenolic compound, data fusion and chemometrics. **Food Chemistry**, v. 215, p. 245-255, 2017.
- BALLABIO, D.; CONSONNI, V. Classification tools in chemistry. Part 1: linear models. PLS-DA. **Analytical methods**, v. 5, n. 16, p. 3790-3798, 2013.
- BASTOS, D. E. M.; OLIVEIRA, D. M.; MATSUMOTO, R. L.; CARVALHO, P. O.; RIBEIRO, M. L. Yerba maté: Pharmacological Properties, Research and Biotechnology. **Medicinal and Aromatic Plant Science and Biotechnology**, v. 1, n. 1, p-37-46, 2007.
- BEĆ, K. B.; GRABSKA, J.; HUCK, C. W. Miniaturized NIR spectroscopy in food analysis and quality control: Promises, challenges, and perspectives. **Foods**, v. 11, n. 10, p. 1465-1518, 2022.

BELCHIOR, V.; BOTELHO, B. G.; FRANCA, A. S. Comparison of spectroscopy-based methods and chemometrics to confirm classification of specialty coffees. **Foods**, v. 11, n. 11, p. 1655-1665, 2022.

BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, M. B. **Chemometrics: a practical guide**. New York: Wiley-Interscience, 1998.

BENDLIN, R. de C. da S. **Secagem Convectiva De Erva-Mate Secagem Convectiva De Erva-Mate**. 93 f. Dissertação (Mestrado em Engenharia de Alimentos) - Departamento de Engenharia Química e de Alimentos, Universidade Federal de Santa Catarina, Florianópolis, 2003.

BORGES, L. R.; LAZZARI, S. M.N.; LAZZARI, F. A. Análise de matérias estranhas em amostras de erva-mate, *Ilex paraguariensis* St. Hil., provenientes de sistemas de cultivo nativo e adensado. **Revista do Instituto Adolfo Lutz**, v. 62, n. 2, p. 77-82, 2003.

BRAGHINI, F.; GIANE DE CARLI, C.; BONSAGLIA, B.; et al. Composição físico-química de Erva-mate, Antes e Após Simulação do Chimarrão. **Pesquisa Agropecuária Gaúcha**, v. 20, n. 1/2, p. 7-15, 2014.

BRASIL. **Resolução RDC N° 716, de 01 de julho de 2022**, Agência Nacional de Vigilância Sanitária, Requisitos sanitários do café, cevada, chás, erva-mate, especiarias, temperos e molhos. Diário Oficial [da] República Federativa do Brasil, Brasília, DF, 01 jul. 2022.

BRERETON, R. G. Pattern recognition in chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 149, p. 90-96, 2015.

BRUICE, P. Y. **Organic chemistry**. Upper Saddle River, NJ: Pearson Prentice Hall, 2007.

CADET, F. Measurement of sugar content by multidimensional analysis and mid-infrared spectroscopy. **Talanta**, v. 48, n. 4, p. 867-875, 1999.

CARRERA, M.; GALLARDO, J. M. Determination of the geographical origin of all commercial hake species by stable isotope ratio (SIR) analysis. **Journal of Agricultural and Food Chemistry**, v. 65, n. 5, p. 1070-1077, 2017.

CARVALHO, I. DE S. B. **Saponificação de ácidos graxos e triacilglicerídeos: Avaliação por espectrometria de infravermelho**. 53 f. Monografia (Especialização em Engenharia de Recursos Minerais) - Departamento de Engenharia de Minas, Universidade Federal de Minas Gerais, Belo Horizonte, 2013.

CEPEA. Centro de Estudos Avançados em Economia Aplicada. **PIB do Agronegócio Brasileiro**. 2023. Disponível em: <https://www.cepea.esalq.usp.br/br/pib-do-agronegocio-brasileiro.aspx>. Acesso em: 3 de set. de 2023.

CHAKRAVARTULA, S. S. N. 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, p. 108816-108827, 2022.

CHATFIELD, C.; COLLINS, A. J. **Introduction to multivariate analysis**. Boston: Springer, 2013.

CHRISTOU, C.; AGAPIOU, A.; KOKKINOFTA, R. Use of FTIR spectroscopy and chemometrics for the classification of carobs origin. **Journal of Advanced Research**, v. 10, p. 1-8, 2018.

CORREIA, P. R. M.; FERREIRA, M. M. C. Reconhecimento de padrões por métodos não supervisionados: Explorando procedimentos quimiométricos para tratamento de dados analíticos. **Química Nova**, v. 30, n. 2, p. 481–487, 2007.

CORSARO, C. et al. HR-MAS and NMR towards foodomics. **Food Research International**, v. 89, p. 1085-1094, 2016.

COSTA FILHO, P. A. da; POPPI, R. J. Aplicação de algoritmos genéticos na seleção de variáveis em espectroscopia no infravermelho médio: determinação simultânea de glicose, maltose e frutose. **Química Nova**, v. 25, p. 46-52, 2002.

COZZOLINO, D.; RESTAINO, E.; FASSIO, A. Discrimination of yerba mate (*Ilex paraguayensis* St. Hil.) samples according to their geographical origin by means of near infrared spectroscopy and multivariate analysis. **Sensing and Instrumentation for Food Quality and Safety**, v. 4, p. 67-72, 2010.

CRAVO FILHO, R. F.; NAVAS, R.; GONÇALVES, E. M. Características físico-químicas e fenóis totais em frutos de juçara em diferentes condições ambientais. **Revista Agro@Mambiente on-Line**, v. 11, n. 4, p. 331-335, 2017.

CROGE, C. P.; CUQUEL, F. L.; PINTRO, P. T. M. Yerba mate: cultivation systems, processing and chemical composition. A review. **Scientia Agricola**, v. 78, e20190259, 2020.

DALLAGO, R. M. et al. Analysis of volatile compounds of *Ilex paraguariensis* A. St.-Hil. and its main adulterating species. **Ciência e Agrotecnologia**, v. 35, p. 1166-1171, 2011.

DAMASCENO, D.; TOLEDO, T. G.; GODINHO, M. S.; et al. Análise multivariada de imagens na química: Um experimento para determinação do pH de águas potáveis. **Química Nova**, v. 38, n. 6, p. 836–841, 2015.

DANIEL, O. **Erva-mate: sistema de produção e processamento industrial**. 1. ed. Dourados: Ed. da UFGD, 2009.

DANTAS, K. C. F. **Desenvolvimento de método quimiométrico baseado em análise de imagens digitais para a quantificação dos corantes amarelo crepúsculo e tartrazina em bebidas**. 109 f. Dissertação (Mestrado em Química) - Departamento de Química, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

DARTORA, N. et al. UPLC-PDA–MS evaluation of bioactive compounds from leaves of *Ilex paraguariensis* with different growth conditions, treatments and ageing. **Food Chemistry**, v. 129, n. 4, p. 1453-1461, 2011.

DA SILVA, W.; ROSAS, C. A. Da erva-mate aos ervateiros: uma análise de sua dinâmica territorial no município de São Mateus do Sul, PR. **Terr@ Plural**, v. 16, p. 1-16, 2022.

DA SILVEIRA, T. F. F.; MEINHART, A. D.; DE SOUZA, T. C. L.; TEIXEIRA FILHO, J.; GODOY, H. T. Phenolic compounds from yerba mate based beverages - A multivariate optimisation. **Food Chemistry**, v. 190, p. 1159–1167, 2016.

DE ALMEIDA, A. G. et al. Direct NIR spectral determination of genetic improvement, light availability, and their interaction effects on chemically selected yerba-mate leaves. **Microchemical Journal**, v. 191, p. 108828-108835 2023.

DE ARAGÃO, N. M. et al. Efeito da acidez e de modificadores orgânicos na determinação de metilxantinas: um experimento de cromatografia líquida de alta eficiência (CLAE) empregando otimização uni e multivariada. **Química Nova**, v. 32, p. 2482-2486, 2009.

DE LUCA, S.; CIOTOLI, E.; BIANCOLILLO, A.; et al. Simultaneous quantification of caffeine and chlorogenic acid in coffee green beans and varietal classification of the samples by HPLC-DAD coupled with chemometrics. **Environmental Science and Pollution Research**, v. 25, p. 28748–28759, 2018.

DE OLIVEIRA, L. H.; MARCA, T. **Estudo da composição química da erva-mate: comparação entre a folha da erva-mate sob sombreamento e intensa luminosidade**. Anais do IV Simpósio das Engenharias ACET, p. 39, 2020.

DE OLIVEIRA, S. V.; WAQUIL, P. D.. Dinâmica de produção e comercialização da erva-mate no Rio Grande do Sul, Brasil. **Ciência Rural**, v. 45, n. 4, p. 750-756, 2015.

DE OLIVEIRA, Y.; ROTTA, E. Área de distribuição natural de erva-mate (*Ilex paraguariensis* St. Hil.). **Unidade Regional de Pesquisa Florestal Centro-Sul**. 1985.

DERAL. Departamento de Economia Rural. **Prognóstico Agropecuário Erva-mate**, v.13, n.39, p.1-9, 2021.

DESPAGNE, F.; MASSART, D.. Variable selection for neural networks in multivariate calibration. **Chemometrics and intelligent laboratory systems**, v. 40, n. 2, p. 145-163, 1998.

DE SOUZA, J. S.; FERRÃO, M. F.. Aplicações da espectroscopia no infravermelho no controle de qualidade de medicamentos contendo diclofenaco de potássio. Parte I: Dosagem por regressão multivariada. **Revista Brasileira de Ciências Farmacêuticas**, v. 42, p. 437-445, 2006.

DIAS, S. L. P. et al. **Química Analítica: Teoria e Prática Essenciais**. Porto Alegre: Bookman Editora, 2016.

DIGGELMANN, P. O caso da quinoa: reflexões sobre globalização alimentar. **Revista Mangút: Conexões Gastronômicas**, v. 1, n. 2, p.72-83, 2021.

DINIZ, P. H. G. Chemometrics-assisted color histogram-based analytical systems. **Journal of Chemometrics**, v. 34, n. 12, p. 1-27, 2020.

DINIZ, P. H. G. **Novas estratégias para classificação simultânea do tipo e origem geográfica de chás**. 148 f. Tese (Doutorado em Química) - Departamento de Química, Universidade Federal da Paraíba, João Pessoa, 2013.

DJOKAM, M. et al. Hyperspectral imaging as a rapid quality control method for herbal tea blends. **Applied Sciences**, v. 7, n. 3, p. 1-16, 2017.

DOĞAN, C. et al. Non-targeted approach to detect pistachio authenticity based on digital image processing and hybrid machine learning model. **Journal of Food Measurement and Characterization**, v. 17, n. 2, p. 1693-1702, 2023.

DORTZBACH, D. et al. Influência do meio geográfico nas características do produto erva-mate. **Research, Society and Development**, v. 9, n. 10, p. e7489109165-e7489109165, 2020.

DOS SANTOS, H. D.; ALVARENGA, Yasmin Almeida; BOFFO, Elisangela Fabiana. ¹H NMR metabolic fingerprinting of Chapada Diamantina/Bahia (Brazil) coffees as a tool to assessing their qualities. **Microchemical Journal**, v. 152, p. 104293-104300, 2020.

DUARTE, I. F.; BARROS, A.; DELGADILLO, I.; ALMEIDA, C.; GIL, A. M. Application of FTIR spectroscopy for the quantification of sugars in mango as a function of ripening. **Journal of Agricultural and Food Chemistry**, v. 50, n. 11, p. 3104–3111, 2002.

ERCIOGLU, E.; VELIOGLU, H. M.; BOYACI, I. H. Determination of terpenoid contents of aromatic plants using NIRS. **Talanta**, v. 178, p. 716-721, 2018.

FABER, J.; FONSECA, Lilian Martins. How sample size influences research outcomes. **Dental press journal of orthodontics**, v. 19, p. 27-29, 2014.

FABER, N. M.; RAJKÓ, R. How to avoid over-fitting in multivariate calibration—The conventional validation approach and an alternative. **Analytica Chimica Acta**, v. 595, n. 1-2, p. 98-106, 2007.

FENOGLIO, D. et al. Active food additive based on encapsulated yerba mate (*Ilex paraguariensis*) extract: effect of drying methods on the oxidative stability of a real food matrix (mayonnaise). **Journal of Food Science and Technology**, v. 58, p. 1574-1584, 2021.

FERMO, I. R. et al. Development of a low-cost digital image processing system for oranges selection using hopfield networks. **Food and bioproducts processing**, v. 125, p. 181-192, 2021.

FERNANDES, C. E. F.; SCAPINELLO, J.; BOHN, A.; et al. Phytochemical profile, antioxidant and antimicrobial activity of extracts obtained from erva-mate (*Ilex paraguariensis*) fruit using compressed propane and supercritical CO₂. **Journal of Food Science and Technology**, v. 54, n. 1, p. 98–104, 2017.

FERNANDES, A. P. **Emprego de estratégias quimiométricas para a diferenciação de alimentos de acordo com as formas de produção**. 228 f. Tese (Doutorado em Ciências) - Departamento de Química, Universidade Federal de São Carlos, 2006.

FERRÃO, M. F.; MELLO, C.; BORIN, A.; MARETTO, D. A.; POPPI, R. J. LS-SVM: Uma nova ferramenta quimiométrica para regressão multivariada. Comparação de modelos de regressão LS-SVM e PLS na quantificação de adulterantes em leite em pó empregando NIR. **Química Nova**, v. 30, n. 4, p. 852–859, 2007.

FERRÃO, M. F. Técnicas de reflexão no infravermelho aplicadas na análise de alimentos. **Tecno-Lógica**, v. 5, n. September, p. 65–85, 2001.

FERREIRA, M. M. C. **Quimiometria: Conceitos, Métodos e Aplicações**. 1 ed. Campinas: Editora da Unicamp, 2015a. 496 p.

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

FERRERA, T. S. et al. Substâncias fenólicas, flavonoides e capacidade antioxidante em erva-mate sob diferentes coberturas do solo e sombreamentos. **Revista Brasileira de Plantas Mediciniais**, v. 18, p. 588-596, 2016.

FIANI, B. et al. The neurophysiology of caffeine as a central nervous system stimulant and the resultant effects on cognitive function. **Cureus**, v. 13, n. 5, e15032, 2021.

FIGUEIRA, L. C. **Espectroscopia Vibracional (No Infravermelho E Raman) E Espectrofotometria De Absorção Uv-Vis Dos Óleos De Buriti (Mauritia Flexuosa) E De Pataua (Oenocarpus Bataua)**. 72 f. Dissertação (Mestrado em Recursos Naturais) - Departamento de Recursos Naturais, Universidade Federal do Oeste do Pará, 2012.

FILIP, R. et al. Antioxidant activity of *Ilex paraguariensis* and related species. **Nutrition research**, v. 20, n. 10, p. 1437-1446, 2000.

FREITAS, D. S.; NUNES, W. S.; APPARECIDO, R. P.; LOPES, T. I. B.; ALCANTARA, G. B. NMR-based approach reveals seasonal metabolic changes in mate (*Ilex paraguariensis* A. St.-Hil.). **Magnetic Resonance in Chemistry**, v. 56, p. 311-320, 2018.

FRIZON, C. N. T.; OLIVEIRA, G. A.; PERUSSELLO, C. A.; et al. Determination of total phenolic compounds in yerba mate (*Ilex paraguariensis*) combining near infrared spectroscopy (NIR) and multivariate analysis. **LWT - Food Science and Technology**, v. 60, n. 2, p. 795–801, 2015.

GELADI, P. Chemometrics in spectroscopy. Part 1. Classical chemometrics. **Spectrochimica Acta Part B: Atomic Spectroscopy**, v. 58, n. 5, p. 767–782, 2003.

GELADI, P.; WOLD, S.; ESBENSEN, K. Image analysis and chemical information in images. **Analytica Chimica Acta**, v. 191, p. 473–480, 1986.

GERBER, T. et al. Yerba mate (*Ilex paraguariensis* A. St.-Hil.) for new therapeutic and nutraceutical interventions: A review of patents issued in the last 20 years (2000–2020). **Phytotherapy Research**, v. 37, n. 2, p. 527-548, 2023.

GOGIASHVILI, M. et al. HR-MAS NMR based quantitative metabolomics in breast cancer. **Metabolites**, v. 9, n. 2, p. 1-27, 2019.

GOLEMANOV, K. et al. The role of the hydrophobic phase in the unique rheological properties of saponin adsorption layers. **Soft matter**, v. 10, n. 36, p. 7034-7044, 2014.

GOMES, A. et al. The successive projections algorithm for interval selection in PLS. **Microchemical Journal**, v. 110, p. 202-208, 2013.

GONZALEZ, R. C.; WOODS, R. E. **Processamento de imagens digitais**. Editora Blucher, 2010.

GOULART, I. C. G. dos R.; JUNIOR, J. F. P. **Erva 20: sistema de produção de erva-mate**. 2019. 154 f. Embrapa Florestas, Brasília, 2019.

HANSEL, F. A. et al. Perspectiva da espectroscopia de infravermelho próximo (nir) na análise sensorial de erva-mate chimarrão em pó. In: **Congreso Sudamericano de la Yerba Mate. 4. Reunión Técnica de la Yerba Mate. 4. Feria de Agronegocios de la Yerba Mate. 2. 2006 11 05-08, 05 al 08 de noviembre de 2006. Posadas, Misiones. AR. 2006.**

HAURESKO, C.; CORREIA, R. L.; GOMES, M. F. V. B. A relação entre a conservação ambiental da floresta com araucárias e os sistemas faxinais no paran . **Revista Pegada**, v. 18, n. 1, p. 131-151, 2017.

IBGE. Produção da Extração Vegetal e da Silvicultura. Rio de Janeiro, IBGE, 2021. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9105-producao-da-extracao-vegetal-e-da-silvicultura.html>. Acesso em: 6 de ago. de 2023.

IG-Mathe. Reconhecimento da IG S o Matheus. 2023. Disponível em: <https://igmathe.com.br/>. Acesso em: 7 de jan. de 2023.

INPI. Instituto Nacional da Propriedade Industrial. **Guia b sico de indica o geogr fica**. Atualizado em 01 fev. 2021. Disponível em: <http://antigo.inpi.gov.br/menu-servicos/indicacao-geografica/pedidos-de-indicacao-geografica-no-brasil>. Acesso em: 21 de jun. 2023.

INPI. Instituto Nacional da Propriedade Industrial. **Indica es Geogr ficas**. Atualizado em: 07 de jul. 2023. Disponível em: <https://www.gov.br/inpi/pt-br/servicos/indicacoes-geograficas>. Acesso em: 21 de jun. 2023.

JEONG, C.-H.; JEONG, H. R.; CHOI, G. N.; KIM, D.-O.; LEE, U.; HEO, H. J. Neuroprotective and anti-oxidant effects of caffeic acid isolated from *Erigeron annuus* leaf. **Chinese Medicine**, v. 6, n. 25, p. 1-9, 2011.

JOLLIFFE, I. T. Principal component analysis and factor analysis. In: **Principal component analysis**. Springer, New York, NY, 1986. p. 115-128.

KAHMANN, A. et al. Near infrared spectroscopy and element concentration analysis for assessing yerba mate (*Ilex paraguariensis*) samples according to the country of origin. **Computers and Electronics in Agriculture**, v. 140, p. 348-360, 2017.

KAISER, C. R. RMN 2D: Detec o inversa e gradiente de campo na determina o estrutural de compostos org nicos. **Qu mica Nova**, v. 23, p. 231-236, 2000.

KEELER, J. **Understanding NMR spectroscopy**. John Wiley & Sons, 2010.

KENNARD, R. W.; STONE, Larry A. Computer aided design of experiments. **Technometrics**, v. 11, n. 1, p. 137-148, 1969.

KHAKSAR, G. et al. Brewing conditions and low-temperature storage affect the total phenolics, caffeoylquinic acid content, and antioxidant activity of yerba mate infusion. **Research Square**, v.1, p. 1-19, 2023.

KHARBACH, M. et al. Current Application of Advancing Spectroscopy Techniques in Food Analysis: Data Handling with Chemometric Approaches. **Foods**, v. 12, n. 14, p. 1-46, 2023.

KOCH, W.; KUKULA-KOCH, W.; KOMSTA, L. Black tea samples origin discrimination using analytical investigations of secondary metabolites, antiradical scavenging activity and chemometric approach. **Molecules**, v. 23, n. 9, p. 1-12, 2018.

KONIECZYNSKI, P.; VIAPIANA, A.; WESOLOWSKI, M. Comparison of Infusions from Black and Green Teas (*Camellia sinensis* L. Kuntze) and Erva-mate (*Ilex paraguariensis* A. St.-Hil.) Based on the Content of Essential Elements, Secondary Metabolites, and Antioxidant Activity. **Food Analytical Methods**, v. 10, n. 9, p. 3063–3070, 2017.

KOTINA, E.; PLOSKIKH, V.; SHIROKOLOBOV, A. Digital Image Processing in Nuclear Medicine. **Physics of Particles and Nuclei**, v. 53, n. 2, p. 535-540, 2022.

KRSTINIĆ, D. et al. Multi-label classifier performance evaluation with confusion matrix. **Computer Science & Information Technology**, v. 1, p. 1-14, 2020.

KUCHARSKA-AMBROŹEJ, K.; KARPINSKA, J. The application of spectroscopic techniques in combination with chemometrics for detection adulteration of some herbs and spices. **Microchemical Journal**, v. 153, p. 104278-104287, 2020.

LI, J.; MA, C; CHEN, W.; et al. Study of the aroma formation and transformation during the manufacturing process of oolong tea by solid-phase micro-extraction and gas chromatography–mass spectrometry combined with chemometrics. **Food Research International**, v. 108, n. March, p. 413–422, 2018.

LI, X. et al. Determination of tea polyphenols content by infrared spectroscopy coupled with iPLS and random frog techniques. **Computers and Electronics in Agriculture**, v. 112, p. 28-35, 2015.

LIMA, P. C. de. **Discriminação de erva-mate para chimarrão quanto à origem geográfica e presença de açúcar utilizando FTIR e quimiometria**. 2019. 39 f. Trabalho de Conclusão de Curso (Engenharia de Alimentos) - Departamento Acadêmico de Alimentos, Universidade Tecnológica Federal do Paraná, Campo Mourão, 2019.

LIU, Y. et al. Rapid identification of the green tea geographical origin and processing month based on near-infrared hyperspectral imaging combined with chemometrics. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 267, s/n, 2022.

LU, Y. et al. Quantitative Assessment of Cold Injury in Tea Plants by Terahertz Spectroscopy Method. **Agronomy**, v. 13, n. 5, p. 1376-1389, 2023.

LUTOMSKI, P.; GOZDZIEWSKA, M.; FLOREK-LUSZCZKI, M.. Health properties of yerba mate. **Annals of Agricultural and Environmental Medicine**, v. 27, n. 2, p. 310-313, 2020.

LUYKX, D. M.; VAN RUTH, S. M. An overview of analytical methods for determining the geographical origin of food products. **Food chemistry**, v. 107, n. 2, p. 897-911, 2008.

MACCARI JR, A. **Análise do pré-processamento da erva-mate para chimarrão**. 2005. 215 f. Tese (Doutorado em Engenharia Agrícola) – Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas, 2005.

MACCARI JR, A. et al. Indústria ervateira no estado do Paraná II–Fornecimento de matéria-prima. **Revista Acadêmica Ciência Animal**, v. 4, n. 1, p. 63-70, 2006.

MAIZTEGUI, B. et al. Dietary Supplementation with Yerba Mate (*Ilex paraguariensis*) Infusion Increases IRS-1 and PI3K mRNA Levels and Enhances Insulin Sensitivity and Secretion in Rat Pancreatic Islets. **Plants**, v. 12, n. 14, p. 2620-2631, 2023.

MAO, Y. et al. Prediction of tea polyphenols, free amino acids and caffeine content in tea leaves during wilting and fermentation using hyperspectral imaging. **Foods**, v. 11, n. 16, p. 2537-2551, 2022.

MAPA. Ministério da Agricultura, Pecuária e Abastecimento. **Portaria nº 573, de 9 de maio de 2022**. Institui o Programa Nacional de Prevenção e Combate à Fraude e Clandestinidade em Produtos de Origem Vegetal (PNFRAUDE). Diário Oficial [da] República Federativa do Brasil, Brasília, DF, 09 mar. 2022.

MARCELO, M. C. A.; POZEBON, D.; FERRÃO, M. F. Authentication of yerba mate according to the country of origin by using Fourier transform infrared (FTIR) associated with chemometrics. **Food Additives & Contaminants: Part A**, v. 32, n. 8, p. 1215-1222, 2015.

MARCELO, M. C. A.; MARTINS, C. A.; POZEBON, D.; DRESSLER, V. L.; FERRÃO, M. F. Classification of yerba mate (*Ilex paraguariensis*) according to the country of origin based on element concentrations. **Microchemical Journal**, v. 117, p. 164–171, 2014.

MARQUES, V.; FARAH, A. Chlorogenic acids and related compounds in medicinal plants and infusions. **Food Chemistry**, v. 113, n. 4, p. 1370-1376, 2009.

MARTINS, A. R. **Desenvolvimento de método para discriminação entre marcas de uísque e identificação de adulteração por espectroscopia UV-Vis e PLS-DA**. 101 f. Dissertação (Mestrado em Química) - Instituto de Química, Universidade de Brasília, Brasília, 2015.

MAZUR, L.; PERALTA-ZAMORA, P. G.; DEMCZUK, B.; RIBANI, R. H. Application of multivariate calibration and NIR spectroscopy for the quantification of methylxanthines in yerba mate (*Ilex paraguariensis*). **Journal of Food Composition and Analysis**, v. 35, n. 2, p. 55–60, 2014.

MAZZEI, P.; PICCOLO, A. HRMAS NMR spectroscopy applications in agriculture. **Chemical and Biological Technologies in Agriculture**, v. 4, n. 1, p. 1-13, 2017.

MENG, W. et al. Geographical origin discrimination of oolong tea (TieGuanYin, *Camellia sinensis* (L.) O. Kuntze) using proton nuclear magnetic resonance spectroscopy and near-infrared spectroscopy. **Food Analytical Methods**, v. 10, p. 3508-3522, 2017.

MEYER, S. et al. Relationships between optically assessed polyphenols and chlorophyll contents, and leaf mass per area ratio in woody plants: a signature of the carbon–nitrogen balance within leaves?. **Plant, Cell & Environment**, v. 29, n. 7, p. 1338-1348, 2006.

MILANEZ, KDT de M. et al. Classification of extra virgin olive oil and verification of adulteration using digital images and discriminant analysis. **Analytical Methods**, v. 7, n. 20, p. 8839-8846, 2015.

MONTEIRO, A. R. D. **Estudo de aspectos estatísticos da espectroscopia na região do infravermelho próximo**. 183 f. Dissertação (Mestrado em Engenharia Química) - Instituto Alberto Luiz Coimbra, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011.

NABECHIMA, G. H. et al. Thermal inactivation of peroxidase and polyphenoloxidase enzymes in mate leaves (*Ilex paraguariensis*) in a conveyor belt oven. **CyTA-Journal of Food**, v. 12, n. 4, p. 399-406, 2014.

NASCIMENTO, C. **Ressonância magnética nuclear**. Editora Blucher, 2016. *E-book*. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9788521210191/>>. Acesso em: 10 jul. 2023.

OLIVEIRA, S. C. de et al. **Espectroscopia no infravermelho associada a quimiometria para a determinação de parâmetros de qualidade e de indicação geográfica de cachaças**. 2018. 98 f. Dissertação (Mestrado em Química), Centro de Ciências e Tecnologia, Universidade Estadual da Paraíba, Campina Grande, 2018.

OLIVEIRA, C. S.; CARLOS, E. F.; VIEIRA, L. G. E.; LIÃO, L. M.; ALCANTARA, G. B. HR-MAS NMR metabolomics of ‘swingle’ citrumelo rootstock genetically modified to overproduce proline. **Magnetic Resonance in Chemistry**, v. 52, p. 422-429, 2014.

OTTO, M. **Chemometrics: statistics and computer application in analytical chemistry**. Weinheim: John Wiley & Sons, 2016.

PANZL, M. V. et al. Evaluation of polycyclic aromatic hydrocarbons in dried leaves of Yerba Mate (*Ilex paraguariensis*) and Their Extraction into Infusions. **Polycyclic Aromatic Compounds**, v. 43, n. 2, p. 1575-1589, 2023.

PANZL, M. V. et al. Analysis of polyphenols and xanthines in yerba mate (*Ilex paraguariensis*) infusions by high-pressure extraction and ultra-high performance liquid chromatography. **Applied Food Research**, v. 2, n. 2, p. 100192-100204, 2022.

PAGLIOSA, C. M. et al. Bitterness in yerba mate (*Ilex paraguariensis*) leaves. **Journal of Sensory Studies**, v. 24, n. 3, p. 415-426, 2009.

PARREIRA, T. F. **Utilização de métodos quimiométricos em dados de natureza multivariada**. 106 f. Dissertação (Mestrado em Química) - Instituto de Química, Universidade Estadual de Campinas, Campinas, 2003.

PASQUINI, C. Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. **Journal of the Brazilian Chemical Society**, v. 14, n. 2, p. 198–219, 2003.

PAVIA, D. et al. **Introdução à espectroscopia**. Cengage Learning Brasil, 2016. *E-book*. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788522123391/>. Acesso em: 5 jan. 2022.

PEDRINI, H.; SCHWARTZ, W. R.. **Análise de imagens digitais: princípios, algoritmos e aplicações**. Cengage Learning, 2008.

PEREZ, C. X.; BOTTERON, F.. A practical approach for segmentation of twigs on samples of yerba mate by digital image processing. **IEEE Latin America Transactions**, v. 11, n. 1, p. 21-26, 2013.

PETROSELLI, G. et al. MALDI-and LDI-MS saponin fingerprint of leaves and stick components of commercial yerba mate (*Ilex paraguariensis*). **Journal of Mass Spectrometry**, v. 54, n. 2, p. 195-203, 2019.

PINTO, H. E. et al. Implicações do Selo ARTE para a competitividade de negócios agroalimentares: o caso dos produtos alimentícios artesanais de origem animal. **Research, Society and Development**, v. 9, n. 8, p. e374985352-e374985352, 2020.

PIRES, D. A. de C. K. et al. A Erva-Mate no Planalto Norte Catarinense: os compostos bioativos como variável na determinação das especificidades necessárias ao reconhecimento como Indicação Geográfica. **DRd-Desenvolvimento Regional em debate**, v. 6, n. 2, p. 207-227, 2016.

POLIDORO, A. dos S. et al. Analytical and chemometric strategies for elucidation of yerba mate composition. **Food Chemistry**, v. 429, p. 136918-136928, 2023.

POLIDORO, A. dos S. **Estratégias analíticas e quimiométricas para elucidação da composição química da erva-mate**. 2022. 205 f. Tese (Doutorado em Química) – Instituto de Química, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2022.

POOLE, R. L.; TORDOFF, M. G. The taste of caffeine. **Journal of caffeine research**, v. 7, n. 2, p. 39-52, 2017.

PRABHU, S. et al. Classifications of polyphenols and their potential application in human health and diseases. **Int. J. Physiol. Nutr. Phys. Educ**, v. 6, n. 1, p. 293-301, 2021.

PRESTES, R. A. et al. Caracterização por ressonância magnética nuclear de sucos de maçã obtidos por preparações enzimáticas. **Química Nova**, v. 35, p. 1141-1145, 2012.

QI, D.; MIAO, A.; CAO, J.; et al. Study on the effects of rapid aging technology on the aroma quality of white tea using GC–MS combined with chemometrics: In comparison with natural aged and fresh white tea. **Food Chemistry**, v. 265, p. 189–199, 2018.

RAKOCEVIC, M. et al. Quality of yerba-mate leaves originating from male and female plants. **Pesquisa Florestal Brasileira**, v. 54, p. 71-83, 2007.

REYES-LÓPEZ, L. E. et al. Scaling analysis of an image encryption scheme based on chaotic dynamical systems. **Entropy**, v. 23, n. 6, p. 672, 2021.

RIBEIRO, J. et al. Photoprotector Effect of Emulsions with Yerba-Mate (*Ilex paraguariensis*) Extract. **Scientia Pharmaceutica**, v. 91, n. 2, p. 22, 2023.

RIBEIRO, Eliana P. **Química de alimentos**. Editora Blucher, 2007. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788521215301/>. Acesso em: 5 set. 2022.

ROSSI, G. B.; LOZANO, V. A. Simultaneous determination of quality parameters in yerba mate (*Ilex paraguariensis*) samples by application of near-infrared (NIR) spectroscopy and partial least squares (PLS). **Lwt**, v. 126, p. 109290-109295, 2020.

RZAŞA-DURAN, E. et al. Yerba mate as a source of elements and bioactive compounds with antioxidant activity. **Antioxidants**, v. 11, n. 2, p. 371-384, 2022.

SA, F. A. S.; DE PAULA, J. A. M.; SANTOS, P. A.; OLIVEIRA, L. A. R.; OLIVEIRA, G. A. R.; LIÃO, L. M.; DE PAULA, J. R.; SILVA, M. R. R. Phytochemical analysis and antimicrobial activity of *Myrcia tomentosa* (Aubl.) DC. leaves. **Molecules**, v. 22, n. 1100, p. 1-10, 2017.

SAATH, R. et al. Alterações na composição química e sensorial de café (*Coffea arabica* L.) nos processos pós-colheita. **Energia na Agricultura**, v. 27, n. 2, p. 96-112, 2012.

SAKANE, K. K.; MATSUURA, E. N.; SIMIONI, A. R. Espectroscopia No Infravermelho Como Ferramenta Para Diferenciação De Chá Verde De Agricultura Orgânica E Convencional. **Revista Tecnologia e Tendências**, v. 10, n. 2, p. 59-74, 2019.

SANTAGIULIANA, M. et al. As good as expected? How consumer expectations and addition of vegetable pieces to soups influence sensory perception and liking. **Food & function**, v. 10, n. 2, p. 665-680, 2019.

SANTANA, F. B. de 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, p. 371-381, 2020.

SANTOS, M. de F. C. et al. ¹H HR-MAS NMR chemical profile and chemometric analysis as a tool for quality control of different cultivars of green tea (*Camellia sinensis*). **Food Chemistry**, v. 408, p. 135016, 2023.

SANTOS, M de F. C. **Estudo do perfil químico de cultivares de chá verde (*camellia sinensis*) ao longo da safra e de chás verdes comerciais, através da técnica de RMN HR-MAS de ¹H**. 2019. 173 f. Tese (Doutorado em Química) – Departamento de Química, Universidade Federal do Paraná, Curitiba, 2019.

SANTOS, H. D.; ALVARENGA, Y. A.; BOFFO, E. F. ¹H NMR metabolic fingerprinting of Chapada Diamantina/Bahia (Brazil) coffees as a tool to assessing their qualities. **Microchemical Journal**, v. 152, p. 104293, 2020.

SANTOS, A. D. C. et al. High-resolution magic angle spinning nuclear magnetic resonance in foodstuff analysis. **TrAC Trends in Analytical Chemistry**, v. 73, p. 10-18, 2015.

SCHNEIDER, M.; SCHNEIDER, R. C. S.; CORBELLINI, V. A.; et al. Exploratory Analysis Applied for the Evaluation of Yerba Mate Adulteration (*Ilex paraguariensis*). **Food Analytical Methods**, v. 11, n. 7, p. 2035–2041, 2018.

SCHINDLER, B. et al. Efeito da sazonalidade sobre o rendimento do óleo essencial de *Piper gaudichaudianum* kunth. **Ciencia Florestal**, v. 28, n. 1, p. 263–273, 2018.

SEBRAE. GIESBRECHT, H. O.; MINAS, R. B. A. de; GONÇALVES, M. F. W.; SCHWANKE, F. H. **Indicações geográficas brasileiras – Vinho**. Brasília: Sebrae, 2014.

SEBRAE. GIESBRECHT, H. O.; MINAS, R. B. A. de; GONÇALVES, M. F. W.; SCHWANKE, F. H. **Indicações geográficas brasileiras**. Brasília: Sebrae, 2016.

SHARIFI-ZAHABI, E. et al. A comprehensive insight into the molecular effect of theobromine on cardiovascular-related risk factors: A systematic review of in vitro and in vivo studies. **Phytotherapy Research**, v. 37, n. 9, p. 3765-3779, 2023.

SIGNOR, P.; DORNELLES, R. C.; BAUMEL, A. **Diagnóstico da erva-mate no paraná**. Instituto de Florestas do Paraná, 2016.

SILVERSTEIN, R. et al. **Identificação Espectrométrica de Compostos Orgânicos**. 7 ed. Rio de Janeiro: LTC, 2006.

SKOOG, D. A.; HOLLER, F. J.; CROUCH, S. R. **Principles of Instrumental Analysis**. 6 ed. Belmont, CA: Thomson Brooks/Cole, 2007. 1016 p.

SOLOMON, C.; BRECKON, T. **Fundamentos de processamento digital de imagens: uma abordagem prática com exemplos em MATLAB**. Rio de Janeiro: LTC, 2011.

SOUSA, E. S. et al. **Estratégia para a modelagem de dados de segunda ordem usando cromatografia líquida para a quantificação de pesticidas em hortaliças**. 2020. Tese (Doutorado em Química) – Departamento de Química, Universidade Federal da Paraíba, João Pessoa, 2020.

SOUSA, S. A. A.; MAGALHAES, A.; FERREIRA, M. M. C.. Optimized bucketing for NMR spectra: Three case studies. **Chemometrics and Intelligent Laboratory Systems**, v. 122, p. 93-102, 2013.

SOUZA, A. M.; ET AL. Proposta de tutorial de Quimiometria utilizando técnicas modernas para a análise de solos. In: SIMPÓSIO BRASILEIRO DE EDUCAÇÃO EM SOLOS, 6., Ceará,- 22 a 25 de Maio, 2012. **Resumos Expandidos**.Sobral, p.15, 2012.

SUN, L. et al. Spectral–spatial feature tokenization transformer for hyperspectral image classification. **IEEE Transactions on Geoscience and Remote Sensing**, v. 60, p. 1-14, 2022.

SZAKÁCS, Z.; SÁNTA, Z.. NMR methodological overview. In: **Anthropic awareness**. Elsevier, 2015. p. 257-289.

TAGLANG, R. Understanding Kernel Convolution (Part 2). 05 de out. de 2016. Disponível em: https://taglang.io/blog/post/Understanding_Kernel_Convolution_Part_2/. Acesso em: 20 de jun. de 2023.

TAVARES, L. A.; FERREIRA, A. G.. Análises quali-e quantitativa de cafés comerciais via ressonância magnética nuclear. **Química Nova**, v. 29, p. 911-915, 2006.

TRUZZI, E. et al. Novel application of ¹H NMR spectroscopy coupled with chemometrics for the authentication of dark chocolate. **Food Chemistry**, v. 404, p. 134522-134531, 2023.

TSAGKARIS, A. S. et al. Spelt authenticity assessment using a rapid and simple Fourier transform infrared spectroscopy (FTIR) method combined to advanced chemometrics. **European Food Research and Technology**, v. 249, n. 2, p. 441-450, 2023.

TSUJIMOTO, T.; YOSHITOMI, T.; MARUYAMA, T.; YAMAMOTO, Y. HAKAMATSUKA, T.; UCHIYAMA, N. ¹³C-NMR-based metabolic fingerprinting of *Citrus*-type crude drugs. **Journal of Pharmaceutical and Biomedical Analysis**, v. 161, p. 305-312, 2018.

TUGIZIMANA, F.; PIATER, L.; DUBERY, I. Plant metabolomics: A new frontier in phytochemical analysis. **South African Journal of Science**, v. 109, n. 5-6, p. 01-11, 2013.

VALDUGA, A.T.; FINZER, J.R.D.; MOSELE, S.H. **Processamento de erva-mate**. Erechim: Edifapes, p.184, 2003.

VASCONCELLOS, A.C.; FRAZZON, J.; NOREÑA, C. P. Z. Phenolic compounds present in Yerba mate potentially increase human health: A critical review. **Plant Foods for Human Nutrition**, v. 77, n. 4, p. 495-503, 2022.

VIEIRA, T. F. et al. Chemometric approach using ComDim and PLS-DA for discrimination and classification of commercial yerba mate (*Ilex paraguariensis* St. Hil.). **Food Analytical Methods**, v. 13, p. 97-107, 2020.

VIEIRA, T. F. **Avaliação de erva-mate para chimarrão (*Ilex paraguariensis* St. Hil.) por métodos instrumentais e quimiometria**. 2018. Dissertação (Mestrado em Tecnologia de Alimentos) - Programa De Pós-Graduação Em Tecnologia De Alimentos, Universidade Tecnológica Federal do Paraná, Campo Mourão, 2018.

VISCONTI, L. G. et al. Determination of grated hard cheese adulteration by digital image analysis and multivariate analysis. **International Dairy Journal**, v. 138, p. 105539-105547, 2023.

YANG, C. et al. Detection of volatile organic compounds in adulterated tea using Fourier transform infrared spectroscopy and Proton-transfer-reaction mass spectrometry. **Food Chemistry**, v. 423, p. 136308-136318, 2023.

YANG, J.; LEE, J.. Consumer perception and liking, and sensory characteristics of blended teas. **Food science and biotechnology**, v. 29, p. 63-74, 2020.

YE, G. et al. Image encryption scheme based on blind signature and an improved Lorenz system. **Expert Systems with Applications**, v. 205, p. 117709-117720, 2022.

WANG, B.; MALDONADO-DEVINCCI, A. M.; JIANG, L. Evaluating line-broadening factors on a reference spectrum as a bucketing method for NMR based metabolomics. **Analytical biochemistry**, v. 606, p. 113872-113879, 2020.

WENDLING, I., ARANTES, M. S. T, HELM, C. V. **Método caseiro de secagem de erva-mate para chimarrão, em forno de micro-ondas**. 2022. 7 f. Comunicado Técnico, EMBRAPA, 2022.

WESTAD, Frank; MARINI, Federico. Validation of chemometric models—a tutorial. **Analytica Chimica Acta**, v. 893, p. 14-24, 2015.

WESTPHALEN, D. J. et al. Phytochemical composition of yerba mate leaves (*Ilex paraguariensis*) and its relation with cultivation conditions. **Revista Brasileira de Plantas Mediciniais**, v. 22, p. 99-107, 2022.

WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and intelligent laboratory systems**, v. 58, n. 2, p. 109-130, 2001.

WONG, A., LUCAS-TORRES, C.. High-resolution Magic-angle Spinning (HR-MAS) NMR Spectroscopy. Hector C. Keun. **NMR-based Metabolomics**, Chapter 5 in: New Developments in NMR, The Royal Society of Chemistry, nº14, p.133-150, 2018.

WORKMAN, J.; WEYER, L. **Practical Guide and Spectral Atlas for Interpretive Near-Infrared**. Boca Raton, FL: CRC, 2012.

WU, D. et al. Determining farming methods and geographical origin of chinese rice using NIR combined with chemometrics methods. **Journal of Food Measurement and Characterization**, v. 17, p. 1-14, 2023.