



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

JOABEL FREIRE DIAS JUNIOR

**APLICAÇÃO DA ESPECTROSCOPIA NIR E QUIMIOMETRIA PARA ESTUDOS
DE FENOTIPAGEM E DE GENÓTIPOS DE *Agave* spp.**

**CAMPINA GRANDE
2023**

JOABEL FREIRE DIAS JUNIOR

**APLICAÇÃO DA ESPECTROSCOPIA NIR E QUIMIOMETRIA PARA ESTUDOS
DE FENOTIPAGEM E DE GENÓTIPOS DE *Agave spp.***

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

Área de concentração: Metodologias Analíticas Avançadas.

Orientador: Prof. Dr. Simone da Silva Simões.

Coorientador: Prof. Dr. Everaldo Paulo de Medeiros.

**CAMPINA GRANDE
2023**

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

D541a Dias Junior, Joabel Freire.

Aplicação da espectroscopia NIR e quimiometria para estudos de fenotipagem e de genótipos de *Agave spp.* [manuscrito] / Joabel Freire Dias Junior. - 2023.

77 p. : il. colorido.

Digitado.

Dissertação (Mestrado em Química) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2023.

"Orientação : Profa. Dra. Simone da Silva Simões , Departamento de Química - CCT. "

"Coorientação: Prof. Dr. Everaldo Paulo de Medeiros , Embrapa Algodão"

1. Espectroscopia. 2. Quimiometria. 3. Fibra de sisal. I.

Título

21. ed. CDD 540

JOABEL FREIRE DIAS JUNIOR


APLICAÇÃO DA ESPECTROSCOPIA NIR E QUIMIOMETRIA PARA
ESTUDOS DE FENOTIPAGEM DE GENÓTIPOS DE *Agave* spp.

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


Área de concentração:
Metodologias Analíticas Avançadas.

Aprovada em: 25/07/2023.


BANCA EXAMINADORA

Documento assinado digitalmente
 SIMONE DA SILVA SIMÕES
Data: 16/08/2023 10:13:31-0300
Verifique em <https://validar.iti.gov.br>

Prof^a. Dr^a. Simone da Silva Simões (Orientadora)
Universidade Estadual da Paraíba (UEPB)

 (Proc SEI 21156.002048/2023-53 Doc 9194732)

Prof. Dr. Everaldo Paulo de Medeiros (Coorientador)
Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)

Documento assinado digitalmente
 PAULO HENRIQUE GONCALVES DIAS DINIZ
Data: 02/08/2023 15:39:10-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Paulo Henrique Gonçalves Dias Diniz (Membro interno)
Universidade Federal do Oeste da Bahia (UFOB)



Prof. Dr. Tarcisio Marcos de Souza Gondim (Membro externo)
Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)

À minha família e amigos, pela dedicação,
companheirismo e amizade, DEDICO.

AGRADECIMENTOS

A Deus, pela minha vida e por me permitir ultrapassar todos os obstáculos encontrados ao longo da realização deste trabalho.

A minha família, em especial ao meu pai Joabel e minha mãe Maria de Lourdes, aos meus irmãos Valquíria, Carlos e Neto, aos meus sobrinhos Kauã e David que sempre estiveram ao meu lado, me dando suporte, carinho e confiança.

À Universidade Estadual da Paraíba, ao Programa de Pós-Graduação em Química, à Secretaria de Pós-Graduação.

Aos membros avaliadores da banca, que se prontificaram a julgar e enriquecer este trabalho, fomentando debates e reflexões para o avanço no campo acadêmico.

A orientadora Prof^a. Dr. Simone da Silva Simões por toda discussão científica, orientação, amizade e pela contribuição ímpar para a execução desse trabalho.

Ao meu coorientador Dr. Everaldo Paulo de Medeiros por toda discussão científica, orientação e amizade.

À EMBRAPA pela oportunidade de estágio e realização desse trabalho e aos então membros do Laboratório Avançado de Tecnologia Química (LATECQ) Edijane e Joabson, pela amizade e ajuda sempre que solicitada.

A todos os professores que contribuíram para minha formação profissional.

Ao grupo de estudos sobre Metodologias Analíticas e Quimiometria orientado pela Prof^a. Dr^a. Simone da Silva Simões.

Aos meus amigos de laboratório: Maria Barbosa, Daniel e Sabrina pela ajuda, por todos os momentos de apoio, discussões científicas e pela amizade.

Aos meus amigos de curso, pelo apoio, incentivo e amizade.

A todos os meus amigos, agradeço a amizade e o carinho e todos as pessoas que contribuíram direta ou indiretamente para a realização deste trabalho.

À CAPES pela bolsa de estudos.

RESUMO

A fibra do sisal, *Agave sisalana*, pode ser utilizada para diversos fins incluindo principalmente aplicações industriais e artesanato. No entanto, apenas 3% a 5% do material foliar é aproveitado como fibra, sendo os coprodutos, como a fração líquida, mucilagem e bucha, descartados. Após o beneficiamento das folhas, os resíduos são rapidamente fermentados causando um aumento da acidez na fração líquida, o que inviabiliza a agregação de valor. Para os extratos líquidos, a determinação da acidez é uma variável crítica devido a fermentação, perda de massa, alterações na cor, pH, condutividade. Com o propósito de permitir novas aplicações para a fração líquida do sisal, este trabalho objetivou desenvolver estratégias de aplicação da espectroscopia NIR (do inglês: *Near Infrared*) e quimiometria para a predição da acidez na fração líquida, e classificação de genótipos de *Agave* spp. Para isto foram coletadas 40 amostras obtidas de seis genótipos, considerando sete plantas por genótipos. As folhas das plantas foram beneficiadas e processadas para a coleta do extrato líquido. As determinações físico-químicas foram realizadas nas amostras in natura e após tratamento do extrato com o uso de conservante químico. Foram medidas a condutividade elétrica, o pH e a perda de massa dessas amostras diariamente durante 51 dias. Em seguida, determinou-se a acidez total por método convencional, e a adquiriu-se os espectros NIR. Com base na PCA (análise em componentes principais), foi avaliada a tendência de distinção entre os genótipos estudados. Os resultados da PCA indicaram que os genótipos A, C e F (1, 7 e 15) são diferentes em sua composição. Os modelos de reconhecimento de padrões supervisionados utilizando LDA (análise discriminante linear) e o algoritmo ACO (colônia de formigas) para seleção de variáveis, alcançou uma taxa de classificação correta superior a 90%. Para os modelos construídos para a determinação da acidez, utilizando os espectros VIS-NIR e os algoritmos iPLS e iSPA-PLS alcançou-se uma capacidade preditiva satisfatória, com valores de RMSEP de 0,870 e 0,771 mg NaOH/mL, r_{pred} de 0,873 e 0,771 e REP de 8,408 e 10,765 % para iPLS e iSPA-PLS, respectivamente. Diante do exposto pode-se afirmar que aplicação da tecnologia NIR apresenta para a determinação da acidez e classificação de genótipos de *Agave* spp. uma técnica rápida, eficiente e precisa.

Palavras-Chave: *Agave* spp.; Espectroscopia NIR; Quimiometria; Acidez.

ABSTRACT

The sisal fiber, *Agave sisalana*, can be used for various purposes, including primarily industrial applications and crafts. However, only 3% to 5% of the leaf material is utilized as fiber, while co-products such as liquid fraction, mucilage, and pulp are discarded. After leaf processing, the residues undergo rapid fermentation, causing an increase in acidity in the liquid fraction, which hinders value addition. For the liquid extracts, acidity determination is a critical variable due to fermentation, mass loss, color changes, pH, and conductivity alterations. With the aim of enabling new applications for the sisal liquid fraction, this study aimed to develop strategies for applying Near Infrared (NIR) spectroscopy and chemometrics for predicting acidity in the liquid fraction and classifying *Agave* spp. genotypes. To achieve this, 40 samples were collected from six genotypes, with seven plants per genotype. The plant leaves were processed to obtain the liquid extract. Physical and chemical determinations were performed on the samples in their natural state and after treating the extract with a chemical preservative. Electric conductivity, pH, and mass loss were measured daily for 51 days. Subsequently, total acidity was determined through conventional methods, and NIR spectra were acquired. Based on Principal Component Analysis (PCA), the trend of distinction among the studied genotypes was evaluated. PCA results indicated that genotypes A, C, and F (1, 7, and 15) have distinct compositions. Supervised pattern recognition models using Linear Discriminant Analysis (LDA) and the Ant Colony Optimization (ACO) algorithm for variable selection achieved a classification rate exceeding 90%. For the models constructed to determine acidity using VIS-NIR spectra and the iPLS and iSPA-PLS algorithms, a satisfactory predictive capacity was attained, with Root Mean Square Error of Prediction (RMSEP) values of 0.870 and 0.771 mg NaOH/mL, prediction correlation coefficients (r_{pred}) of 0.873 and 0.771, and Relative Error in Prediction (REP) of 8.408 and 10.765% for iPLS and iSPA-PLS, respectively. In light of the above, it can be affirmed that the application of NIR technology presents a fast, efficient, and accurate technique for acidity determination and classification of *Agave* spp. genotypes.

Keywords: *Agave* spp.; NIR spectroscopy; Chemometrics; Acidity.

LISTA DE ILUSTRAÇÕES

Figura 1 – (a) - <i>Agave sisalana</i>. (b) - <i>Agave híbrido</i> 11648.	21
Figura 2 - Diferentes genótipos de <i>Agave</i> spp. do Banco Ativo de germoplasma de da EMBRAPA Algodão.	22
Figura 3 - Diferentes tipos de modo vibracionais para uma molécula triatômica heteronuclear (a) - Vibrações de estiramento, (b) - Vibrações de deformação angular.	26
Figura 4 - (a) - Diagrama de energia potencial para os osciladores harmônicos, (b) – Anarmônico.	27
Figura 5 - Modos de medidas na região NIR. (a) transmitância, (b) transflectância e (c) refletância difusa.	28
Figura 6 - Representação gráfica da projeção das amostras no subespaço A gerado pelas duas componentes principais, PC1 e PC2. t_i é o vetor de escores definido no plano das PC; t_{i1} e t_{i2} são as coordenadas (os escores) de i nesse plano e e_i é o vetor de resíduos, ortogonal ao plano A.	31
Figura 7 - Organização dos dados da calibração multivariada.	34
Figura 8 - Decomposição das matrizes X e Y no algoritmo do PLS.	36
Figura 9 - Sequência de projeções realizadas pelo SPA com $K_c = 3$ e $j = 5$. Resultado da primeira (a) e segunda (b). Neste exemplo, o conjunto de variáveis selecionadas inicia com x_1 e deverá ser $\{x_1, x_2, x_5\}$.	38
Figura 10 - Sistema de medidas espectrais NIR para medidas e análise da estabilidade/degradação do extrato de <i>Agave</i> spp. (a) - Espectrômetro NIR. (b) - Cubeta de quartzo. (c) - Caminho óptico de 50 mm. (d) - Amostras do extrato de <i>Agave</i> spp.	44
Figura 11 - Taxa de degradação para os seis genótipos de <i>Agave</i> spp. (a) A, (b) B, (c) C, (d) D, (e) E e (f) F no período de 51 dias, em função da dose do conservante.	49
Figura 12 - Variação (a) comprimento (T em cm), (b) massa (M em g) das folhas (c) teor de cinzas (CZEL%) e (d) sólidos totais (ST%) no extrato e	50
Figura 13 - Teste de Tukey. (a) CZEL: Cinzas Extrato Líquido, (b) ST: Sólidos Totais, (c) TA: Teor Água, (d) AT: Acidez Titulável, (e) T: Tamanho médio e (f) M: Massa foliar média.	53

Figura 14 - Espectros brutos dos genótipos de <i>Agave</i> spp. completos com faixa removida destacada em vermelho.	54
Figura 15 - (a) faixa de trabalho, (b) dados pré-processados com SNV e (c) 1ª derivada com polinômio de 1ª ordem e (d) 2ª derivada com polinômio de 2ª ordem.	55
Figura 16 - Gráfico de escores para amostras de <i>Agave</i> spp. na região VIS-NIR. (a) todas os genótipos e (b) para os genótipos 1, 7 e 15 (A, C e F).....	56
Figura 17 - (a) Modelo ACO-LDA com SNV para as 5 classes de genótipos de <i>Agave</i> spp. (b) Variáveis selecionadas pelo algoritmo ACO para as 5 classes de genótipos. (c) Modelo ACO-LDA com SNV para 3 classes de genótipos. (d) Variáveis selecionadas pelo algoritmo ACO para 3 classes de genótipos.....	58
Figura 18 - Gráficos de Loadings do modelo (a) PLS as variáveis selecionadas para os modelos (b) iPLS e (c) iSPA-PLS, utilizando 2ª Derivada, com janela de 5 pontos e polinômio de 2º para o PLS e iSPA-PLS e 1ª Derivada, com janela de 7 pontos e polinômio de 1º para o iPLS.	61
Figura 19 - Gráficos de valor predito versus referência obtido para a determinação da acidez pelos modelos (a) PLS e (c) iPLS (e) iSPA-PLS. Gráficos de EJCR com as elipses de confiança contendo seus respectivos pontos teóricos ideais obtidos para a determinação da acidez pelos modelos (b) PLS (d) iPLS e (f) iSPA-PLS, utilizando 2ª Derivada, com janela de 5 pontos e polinômio de 2º para o PLS e iSPA-PLS e 1ª Derivada, com janela de 7 pontos e polinômio de 1º para o iPLS.	63

LISTA DE TABELAS

Tabela 1 - Amostras de genótipos de <i>Agave</i> spp. do Banco de Germoplasma de <i>Agave</i> da EMBRAPA com plantas em idade de 51 meses de cultivada na Estação Experimental da EMBRAPA em Monteiro-PB utilizadas na análise.	41
Tabela 2 - Genótipos estatisticamente diferentes de acordo com o teste Tukey.	52
Tabela 3 - Figuras de mérito para o modelo ACO-LDA para as 5 classes de genótipos de <i>Agave</i> spp.	57
Tabela 4 - Figuras de mérito para o modelo ACO-LDA para as 3 classes de genótipos de <i>Agave</i> spp.	57
Tabela 5 - Resultados dos modelos PLS, iSPA-PLS e iPLS obtidos para determinação da acidez dos genótipos de <i>Agave</i> spp.	60

LISTA DE ABREVIATURAS E SIGLAS

ACO	Algoritmo de Colônia de Formigas
AT	Acidez Titulável
CV	Validação Cruzada
CZ%	Teor de Cinzas
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
EJCR	Região Elíptica de Confiança Conjunta
FAO	Organização das Nações Unidas para a agricultura e a Alimentação
FAR	Infravermelho Distante
FT-IR	Espectroscopia com Transformada de Fourier
HPAEC-PAD	Cromatografia de Troca Aniônica de Alto Desempenho com Detecção Amperométrica Pulsada
IBGE	IBGE – Instituto Brasileiro de Geografia e Estatística
IDH	IDH – Índice de Desenvolvimento Humano
iPLS	Regressão por Mínimos Quadrados Parciais em intervalos
iSPA	Algoritmo de Projeções Sucessivas em intervalos
LATECQ	Laboratório Avançado de Tecnologia Química
LabSIQ	Laboratório de Síntese Inorgânica e Quimiometria
LAD	Análise Discriminante Linear
LV	Variáveis Latentes
MIR	Infravermelho Médio
MLR	Regressão Linear Múltipla
NIR	Infravermelho Próximo
RMN	Ressonância Magnética Nuclear
PC	Componente Principal
PCA	Análise de Componentes Principais

PET	Polietileno Tereftalato
PLS	Mínimos Quadrados Parciais
PLSR	Regressão por Mínimos Quadrados Parciais
R ²	Coeficiente de determinação
Rcv	Coeficiente de correlação de validação cruzada
REP	Erro Relativo de Predição
RMSEP	Raiz Quadrada Média do Erro de Predição
RMSEV	Raiz do Erro Quadrático Médio da Validação
RMSECV	Raiz do Erro Quadrático Médio da Validação Cruzada
RPD	Razão de Desempenho para Desvio
RT	Tempo De Retenção
TA%	Teor de Água
TLC	Cromatografia de Camada Fina
SNV	Varição Normal Padrão
ST	Sólidos Totais
UV	Ultra Violeta

SUMÁRIO

1. INTRODUÇÃO	14
1.1. OBJETIVOS	17
1.1.1. Objetivo Geral	17
1.1.2. Objetivos específicos	17
2. FUNDAMENTAÇÃO TEÓRICA	18
2.1. <i>Agave</i> spp.	18
2.2. Característica Botânica	19
2.2.1. Condições Edafoclimáticas	21
2.3. Aplicações de tecnologias espectroscópicas em matrizes de <i>Agave</i> spp.	23
2.5. Espectroscopia no Infravermelho	24
2.5.1. Espectroscopia no Infravermelho Próximo (NIR)	27
2.6. Quimiometria	28
2.6.1. Análise de Componentes Principais (PCA)	29
2.6.2. Algoritmo de Colônia de Formiga acoplado a Análise Discriminante Linear (ACO-LDA)	31
2.6.3. Calibração Multivariada	33
2.6.3.1. Regressão por Mínimos Quadrados Parciais (PLS)	34
2.6.4. Regressão por Mínimos Quadrados Parciais em intervalos (iPLS)	37
2.6.5. Algoritmo das Projeções Sucessivas em intervalos para Regressão PLS (iSPA-PLS).....	37
3. MATERIAIS E MÉTODOS	40
3.1. Local dos ensaios	40
3.2. Coleta de folhas de <i>Agave</i> spp. e extração da fração líquida	40
3.3. Caracterização físico-química dos extratos processados de <i>Agave</i> spp.	41
3.3.1. Preparo das amostras para a análise de degradação	41

3.3.2. Determinação da perda de massa dos extratos líquidos	41
3.3.3. Determinação de condutividade elétrica em extratos líquidos	42
3.3.4. Determinação de pH em extratos líquidos	42
3.3.5. Determinação de acidez total	42
3.3.6. Determinação do teor de água	42
3.3.7. Determinação de sólidos totais no extrato líquido pela técnica gravimétrico	43
3.3.8. Medidas na região do Infravermelho Próximo - NIR.....	43
3.4. Desenvolvimento dos Modelos Quimiométricos	44
3.4.1. Modelos de classificação (ACO-LDA).....	45
3.4.2. Modelos de calibração (PLS, iPLS, iSPA-PLS).....	45
3.4.3. Softwares.....	47
4. RESULTADOS E DISCUSSÃO	48
4.1. Caracterização físico-química dos extratos processados de <i>Agave</i> spp.	48
4.2. Modelos baseados na espectroscopia VIS-NIR	53
4.2.1. Reconhecimento de Padrões.....	54
4.2.2. Modelos de Regressão (PLS) e Seleção de Variáveis (iSPA PLS e iPLS)	58
5. CONCLUSÕES	65
6. PERSPECTIVAS.....	66
REFERÊNCIAS.....	67
APÊNDICES	76

1. INTRODUÇÃO

A *Agave sisalana*, também conhecido como sisal, é uma planta nativa da Península de Yucatán, no sudoeste do México, mas é encontrada em diversos países, devido à sua boa adaptação em ambientes semiáridos e áridos (TREJO-TORRES, 2018). O Brasil é o maior produtor de sisal do mundo, segundo a Organização das Nações Unidas para a Alimentação e Agricultura (FAO), em 2022 a produção chegou a cerca de 120 mil toneladas de fibra, o que corresponde a 40% de toda a produção mundial (FAO, 2023). O estado da Bahia detém o primeiro lugar na produção, com 94,2%, seguido da Paraíba (5,6%), Ceará (0,1%) e Rio Grande do Norte (0,1%) (EMBRAPA, 2023; IBGE, 2020; SANTOS, SILVA, 2017).

Essa produção está localizada em regiões que apresentam climas semiáridos, os quais não são favoráveis economicamente para o cultivo de outras espécies, como grãos. Dessa forma, o sisal tornou-se uma das poucas atividades agrícolas como fonte de emprego e renda em áreas de cultivo que possuem os menores Índices de Desenvolvimento Humano (IDH = 0,589), gerando cerca de 700 mil empregos diretos ou indiretos em todo o Nordeste, por meio do desfibramento de suas folhas (ANDRADE, ORNELAS, BRANDÃO, 2011).

No processo de extração da folha dura do sisal, obtém-se como principal produto uma fibra rígida, no qual corresponde de 3 a 5% da biomassa das suas folhas, além de coprodutos como mucilagem (15%), bucha composta de fibras curtas (1%) e uma expressiva fração líquida (80%), conhecida como suco de sisal, são produzidos. Dentre os possíveis usos domésticos e industriais da fibra do sisal, inclui-se sua utilização para melhorar a resistência de materiais na construção civil, obtenção de compósitos e indústria têxtil (tapetes, carpetes, capachos e cordas) (ABEBAYEHU, ENGIDA, 2021; DIAS *et al.*, 2021; EMBRAPA, 2023; FERNANDES, 2017; GONDIM *et al.*, 2022; SANTOS, FONTE, LIMA, 2017). No entanto, além da fibra, seus coprodutos que representam de 95 a 97% da massa da folha são descartados ou possuem baixa agregação de valor para outros fins pouco transformados (GONDIM, *et al.*, 2022).

O potencial que os coprodutos obtidos pelo desfibramento do sisal podem gerar, tem direcionado novas aplicações, como, por exemplo, Souza (2017) cita que a mucilagem pode ser utilizada como complemento para alimentação de caprinos e

bovinos. Enquanto, a fração líquida ou suco de sisal podem ser utilizados para o controle de *Aedes aegypti* e pragas agrícolas (BRAGA, V. A. *et al.*, 2018; GONDIM, *et al.*, 2022; OLIVEIRA *et al.*, 2016; SANTOS *et al.*, 2022), sobre *Meloidogyne javanica* em plantas de alface (DAMASCENO, *et al.* 2017), de *Aspergillus welwitschiae* e da podridão vermelha (SILVA *et al.*, 2019).

Entretanto, após a extração da fração líquida, a fermentação natural por microrganismos do próprio ambiente, ocorre de maneira bastante acelerada, atingindo seus níveis máximos nos dois primeiros dias (ZULLO *et al.*, 1984). Várias reações químicas se processam causando aumento da acidez da fração líquida, mesmo ainda misturada com a mucilagem e resíduos de fibras curtas (buchas) de forma a inviabilizar seu uso e aplicação. Portanto, a acidez é uma das variáveis relevantes para estudos de aplicação e agregação de valor à fração líquida.

Em estudos de agregação de valor a coprodutos ou resíduos da biomassa, muitos ensaios de caracterização e análises químicas são requeridos. Apesar da simplicidade dos métodos clássicos usados para essa finalidade, no que se refere a implementação e execução, há imensas desvantagens quando se considera aspectos de tempo, geração de resíduos, infraestrutura, e mais recentemente limites ambientais no tocante a redução das emissões de carbono e uso racional da água, e para alto número de análises de rotina (INSTITUTO ADOLF LUTZ, 2008). Como alternativa, as técnicas espectrais compõem recursos instrumentais e metodológicos cada vez mais explorados para atender diferentes propósitos analíticos (SILVA, *et al.*, 2019; SOUZA *et al.*, 2013; VILVERT *et al.*, 2023).

Dentre elas, espectroscopia no infravermelho próximo (NIR) (FERREIRA, M., 2015), apresenta vantagens na redução de resíduos de forma ambientalmente amigável, de permitir atingir uma alta frequência analítica, com baixa preparação das amostras, possibilidade de medidas não destrutivas, tem se consolidado para viabilizar medidas em campo com a utilização de equipamentos portáteis, e de permitir medidas multicomponentes simultânea em matrizes agrícolas e em plantas (CUI *et al.*, 2012; SOARES *et al.*, 2017; MARTINS, SARGENTELLI, 2021). Devido à complexidade dos espectros obtidos por NIR, há a necessidade da utilização de ferramentas quimiométricas para interpretar a informação analítica.

A aplicação da espectroscopia NIR combinado com a quimiometria para a avaliação de extratos de plantas com diferentes propriedades fitoquímicas têm sido realizadas com sucesso em relação às vantagens supracitadas (GAUB *et al.*, 2004;

YAO *et al.*, 2022; LÓPEZ-HIDALGO, MENÉNDEZ, JORRIN-NOVO, 2021). Através dessa metodologia, é possível identificar a presença de marcadores espectrais que podem estar associados com alguma característica fenotípica como tamanho das folhas, teor de fibras, umidade, nutrientes, resistência a estresses, dentre outras. O que acelera o processo de seleção de genótipos promissores, além de contribuir para a otimização de práticas agrícolas e monitoramento do cultivo (FERREIRA, D., 2015).

Em plantas do gênero da *Agave* spp. apenas um trabalho na literatura científica foi encontrado utilizando NIR e quimiometria. Desenvolvido por Mellado-Mojica e López (2015) teve como propósito estudar e classificar diferentes xaropes de *Agave* spp. de acordo com seus açúcares naturais. Muito embora, não tenham chegado a resultados satisfatórios.

A espectroscopia NIR ainda não foi empregada para classificação e determinação da acidez de genótipos de *Agave* spp. Dessa forma, o presente trabalho propõe a utilização da espectroscopia NIR como ferramenta analítica simples e rápida para classificação e determinação da acidez de diferentes genótipos de *Agave* spp. utilizando Algoritmo de Colônia de Formiga acoplado a Análise Discriminante Linear (ACO-LDA, do inglês, *Ant Colony Optimization coupled with Linear Discriminant Analysis*) e Mínimos Quadrados Parciais (PLS, do inglês *Partial Least Squares*), além de técnicas de seleção de variáveis como o Algoritmo de Projeções Sucessivas acoplado à Análise Discriminante Linear (SPA-LDA, do inglês *Successive Projections Algorithm-Linear Discriminant Analysis*) e Mínimos Quadrados Parciais em intervalos (iPLS, do inglês *interval Partial Least Squares*).

Em função da importância do *Agave* spp. para o cenário econômico, principalmente para o nordeste brasileiro e o pouco percentual de seu aproveitamento, faz-se necessário o desenvolvimento de um método para classificar e garantir a qualidade do extrato líquido de diferentes genótipos de *Agave* spp. pela determinação da acidez, oriundo do desfibramento das suas folhas.

1.1. OBJETIVOS

1.1.1. Objetivo Geral

Desenvolver métodos analíticos baseados na espectroscopia NIR e quimiometria para diferenciar genótipos de *Agave* spp. e quantificar acidez do extrato líquido.

1.1.2. Objetivos específicos

- Registrar os espectros NIR na faixa de 780 a 2500 nm dos extratos líquidos de diferentes genótipos de *Agave* spp.;
- Desenvolver uma metodologia baseada nos espectros NIR e técnicas de reconhecimento de padrões para classificação dos extratos líquidos de diferentes genótipos de *Agave* spp.;
- Desenvolver um modelo de calibração multivariada para predição da acidez total do extrato líquido de diferentes genótipos de *Agave* spp.;
- Avaliar a perda de massa e acidez dos extratos líquidos de diferentes genótipos de *Agave* spp.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. *Agave* spp.

O primeiro gênero de *Agave* spp. a chegar ao Brasil foi o *Agave sisalana*. Introduzido em 1903 no estado da Bahia, mudas oriundas da Florida (EUA), trazidas pelo agrônomo Horácio Urpia Júnior (ALVES, SANTIAGO, 2006). Em 1911, o agrônomo J. Viana Júnior trouxe para a Paraíba as primeiras mudas de *Agave* spp. (NUNES, 1999 apud SILVA, BELTRÃO, 2013).

Na Bahia, entre as décadas de 1960 e 1970, atingiu-se o auge da área cultivada com *Agave* spp., saindo de 58,9 mil hectares, em 1960, para 268,4 mil hectares em 1970, chegando em uma produção de 280.000 toneladas de fibras no mesmo ano, maior pico já registrado (SANTOS, SILVA, 2017). Na década de 1980, houve um declínio na produção de *Agave* spp., que pode estar relacionado ao surgimento de fibras sintéticas, que apresentavam custos mais acessíveis, culminando em uma queda brusca do preço e da demanda pela fibra do sisal, além do baixo índice de modernização, o que afetou diretamente os produtores da região semiárida nordestina (QUEIROGA, 2021; SANTOS, SILVA, 2017; SOUZA, 2013).

Em 2021, segundo dados do Instituto Brasileiro De Geografia e Estatística (IBGE) na Bahia, cerca de 100,3 mil hectares são cultivados. Já na Paraíba, a produção é distribuída em cerca de 5,5 mil hectares, no Ceará (60 hectares) e Rio Grande do Norte (35 hectares), no qual, cerca de 700 mil pessoas obtêm parte de sua renda a partir dessa atividade (IBGE, 2021; QUEIROGA, 2021).

O processo de cultivo do sisal ocorre de forma manual, iniciando pelo corte das folhas das plantas, seguido do desfibramento realizado em máquinas desfibradoras, conhecidas como “máquinas paraibanas”. Tais máquinas apresentam baixa produtividade com valores médios de 150 a 200kg por turno de 10h/dia, além do desperdício que gira em torno de 20 a 30% de fibra da folha, outro ponto negativo está relacionado aos perigos no manuseio da máquina paraibana, sendo responsável por diversas mutilações, devido à proximidade das mãos com o motor da máquina (EMBRAPA, 2023).

Com o passar do tempo, foram desenvolvidas outras desfibradoras como a Corona e a Brasilana, mas não tiveram grande aceitação dos produtores, devido a problemas relacionadas a necessidade de maior quantidade de água, travamentos e

ao enrolamento das fibras. A máquina desfibradora Faustino VII e VIII são as mais modernas, possuem melhor capacidade de produção e garantem mais segurança aos operadores por possuírem esteiras que conduzem as folhas até o motor. Entretanto, a máquina paraibana continua sendo a opção mais acessível para os produtores devido ao custo e simplicidade (QUEIROGA, 2021; EMBRAPA, 2023).

Após o desfibramento, obtêm-se as fibras que precisam ser lavadas com água para remoção do pó e levadas aos varais para secagem ao sol. Posteriormente, as fibras secas são levadas para o seu beneficiamento nas batedeiras para alisamento e seleção. Após essa etapa, as fibras já estão prontas para a comercialização como matéria-prima seja para exportação, para as indústrias nacionais ou para pequenos artesãos (ALVEZ, SANTIAGO, 2006; SANTOS, SILVA, 2017).

A fibra com alta resistência é utilizada na produção de chapéus, cordas, sacos, bolsas, vassouras e barbantes. Também é utilizado para produzir vários tipos de papéis finos e papel kraft. Possui aplicabilidade na indústria automotiva, de eletrodomésticos, de móveis, na obtenção de compósitos, construção civil entre outras (ALVES, SANTIAGO, 2004 *apud* PROSSIGA, 2006). A sua principal utilização tem sido na produção do fio agrícola Baler Twin (SAVASTANO JUNIOR, PIMENTEL, 2000).

Os subprodutos obtidos do desfibramento do sisal são constituídos, em média, 15% de mucilagem, 1% bucha (fibra curta) e 80% fração líquida (suco ou seiva). A mucilagem e a bucha são separadas através de peneiras rotativas, que devem ser instaladas próximas as máquinas desfibradoras para aproveitar ao máximo os subprodutos gerados. A umidade presente na mucilagem corresponde de 90% a 95% da sua massa, a fração líquida é obtida pela extração da umidade da mucilagem (EMBRAPA, 2023).

2.2. Característica Botânica

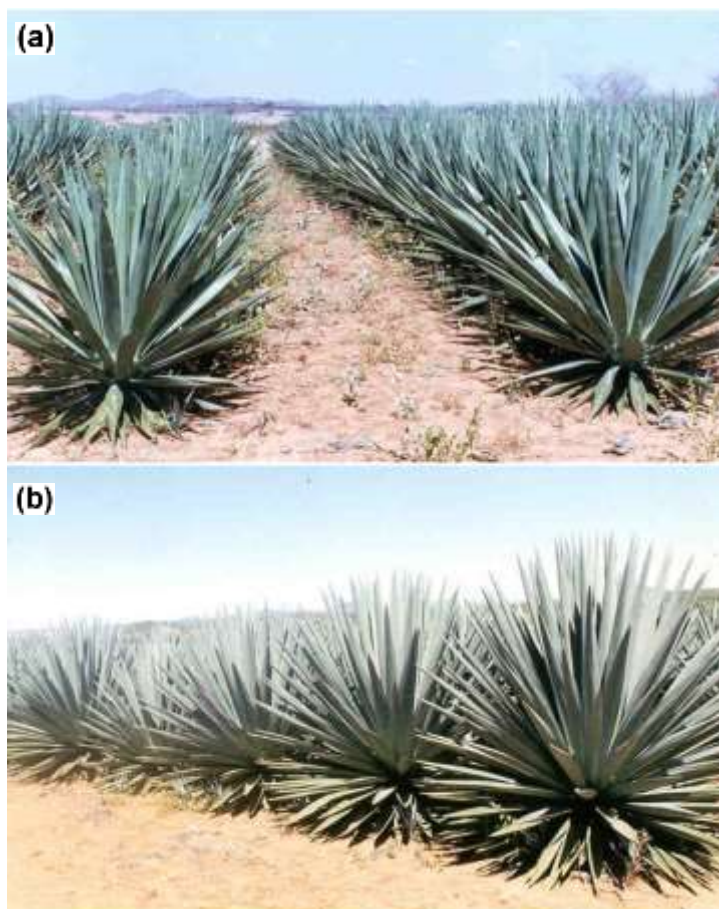
A palavra *Agave* é oriunda do grego “*agauos*” que significa admirável. O gênero *Agave* spp. possui cerca de 300 espécies, sendo 159 localizadas no México (GARCÍA-MENDOZA, MARTÍNEZ, GUTIERREZ, 2019; GSCHAEDLER *et al.*, 2017; MEDINA, 1954). Com base nas características do escapo floral, a *Agave* spp. é classificado no reino *Plantae*, divisão *Magnoliophyta*, classe *Liliopsida*, ordem *Asparagales*, família *Agavaceae* e gênero *Agave* (BARRETO, 2017).

No Brasil a *Agave sisalana* (**Figura 1a**) e a *Agave híbrido 11648* (**Figura 1b**) são os genótipos de maior importância comercial, sendo os mais adaptados e devido a isso, se tornaram os genótipos mais cultivados nas regiões semiáridas do Nordeste (QUEIROGA, 2021; EMBRAPA, 2023).

A *Agave* possui um sistema radicular fasciculado e fibroso, sem uma raiz principal, com dois tipos distintos de raízes: fixadoras e alimentadoras. As raízes fixadoras são ramificadas e podem chegar a mais de três metros de comprimento, responsáveis pela fixação da planta no solo, enquanto as raízes alimentadoras são mais finas e estão inseridas sobre as transportadoras, sendo responsáveis pela absorção de água e nutrientes (MOREIRA *et al.*, 1999).

A planta é acaulescente, o que significa que não possui caule aéreo. Em vez disso, possui um tronco (pseudocaule) ou eixo principal que armazena água e nutrientes e serve como ponto de inserção para as folhas lanceoladas, rígidas e suculentas. As folhas crescem em uma densa roseta, medindo entre 120 e 160 cm de comprimento e 10 a 15 cm de largura na parte média, sem pecíolo. Os bordos das folhas são retos e podem apresentar alguns espinhos rudimentares ou muitos espinhos marginais. Na camada média da folha, os fios das fibras se unem e se lignificam, formando um acúleo terminal de coloração castanho-escuro, com 4 a 5 mm de largura e 20 a 25 mm de comprimento (SILVA *et al.*, 2008).

Figura 1 – (a) - *Agave sisalana*. (b) - *Agave híbrido 11648*.



Fonte: (QUEIROGA, 2021).

2.2.1. Condições Edafoclimáticas

A cultura da *Agave* spp. se desenvolve em regiões semiáridas ou áridas, com baixos índices pluviométricos, solos sílico-arenosos (SUINAGA, SILVA, COUTINHO, 2006). Essas são condições características presentes em proporções típicas do semiárido brasileiro localizando em maior proporção no Nordeste. A adaptação do *Agave* spp. possibilitou o surgimento da atividade econômica sisaleira em grande parte do Nordeste, se tornando uma das poucas opções econômicas para algumas microrregiões carentes (SOARES, ARRUDA, AMARANTE, 2022).

O ambiente ideal para que a planta se desenvolva e cresça bem é em clima quente, com muita luminosidade, e com temperatura média anual entre 20 °C e 28 °C. A precipitação pluviométrica deve estar entre 650 a 1250 mm/ano e solos sílico-arenosos. As medidas médias para a folha de sisal são de 90 cm a 120 cm, chegando a um comprimento máximo de até 2 m, podendo produzir 200 folhas por planta e

chegar a uma altura de 5-6 metros por volta de 7 anos de idade (BROWN, 2002; EMBRAPA, 2023; MARTIN et al., 2009).

O *Agave* híbrido 11648 é o segundo genótipo com maior cultivo no Brasil, sendo originária do Oeste da África e foi obtido através do cruzamento entre *Agave amaniensis* e o *Agave angustifolia* (RAYA, 2018; ROBERT et al., 2008). Em comparação com a *Agave sisalana*, o *Agave* híbrido 11648 é mais resistente a seca e a doença mais relevantes para a cultura, denominada podridão vermelha do tronco, entretanto, apresenta maior suscetibilidade a algumas espécies *Phytophthora* (BAUTISTA-MONTES, HERNÁNDEZ-SORIANO, SIMPSON, 2022). A produção das folhas pode chegar a ser duas ou três vezes maior que o *Agave sisalana*, mas requer um esforço maior do operador da desfibradora (EMBRAPA, 2023; MEDINA, 1954).

As plantas de *Agave* spp. possuem um aspecto morfológico bastante distinto da maioria das plantas. As folhas possuem espinhos e são rígidas, de coloração verde, com largura entre 10 a 15 cm e comprimento de 1 a 2 metros, podem chegar a uma altura de 1,7 a 2 metros (QUEIROGA, 2021), conforme observamos na **Figura 2**. Portanto, não é tarefa fácil identificar e classificar os diferentes genótipos de apenas pelas suas características físicas.

Figura 2 - Diferentes genótipos de *Agave* spp. do Banco Ativo de germoplasma de da EMBRAPA Algodão.



Fonte: (GONDIM et al., 2022).

2.3. Aplicações de tecnologias espectroscópicas em matrizes de *Agave* spp.

Foi desenvolvido um método por Mellado-Mojica e López (2015) para identificação, classificação e discriminação de xaropes de *Agave* spp. a partir de 43 amostras de adoçantes naturais, utilizando espectroscopia MIR e ferramentas quimiométricas, além da cromatografia de troca aniônica de alto desempenho com detecção amperométrica pulsada (HPAEC-PAD). Os espectros MIR foram registrados na faixa entre 2.500 e 15.384 nm com uma resolução nominal de 2.500.000 nm no modo de transmitância (%T). Para a análise quimiométrica, modelos de espectros MIR foram desenvolvidos para carboidratos (8.438–10.526 nm), proteínas (5.814–6.757 nm) e ácidos graxos (3.333–3.521 nm). A análise de componentes principais (PCA) dos espectros MIR permitiu agrupar os adoçantes naturais de acordo com suas origens naturais usando a região de absorção característica para carboidratos. Por HPAEC-PAD os carboidratos foram separados usando um gradiente de acetato de sódio em NaOH 0,15 mol/L em um fluxo de 0,8 mL min⁻¹ e uma temperatura de coluna de 25 °C. Os potenciais aplicados para detecção pelo pulso amperométrico foram E1 (400 ms), E2 (20 ms), E3 (20 ms) e E4 (60 ms) de +0,1, -2,0, +0,6 e -0,1 V, respectivamente. De acordo com os resultados da TLC, o HPAEC-PAD indicou que glicose, frutose e sacarose foram os carboidratos mais abundantes nos xaropes naturais.

SANTOS *et al.*, (2015) desenvolveram um método para isolamento e identificação de produtos químicos da biomassa de *Agave sisalana* por cromatografia de camada fina (TLC), espectroscopia FT-IR, ressonância magnética nuclear (RMN), cromatografia líquida de alta eficiência acoplada ao detector de fotiodo (HPLC-DAD). Os resíduos de sisal (70 kg) foram redissolvidos com água destilada, o extrato filtrado e concentrado rendeu 6,1 kg de extrato bruto. Inicialmente removeu-se por precipitação 83% dos carboidratos. Os espectros de infravermelho do precipitado apresentaram bandas em 2.906 nm (OH), 3349 nm (grupos CH₂) e 6.184 e 6.988 nm (vibrações O=C-O) que indicam a presença principalmente de pectina. Realizado uma extração líquido-líquido com acetato de etila, seguido da remoção dos cristais formados (manitol). A partir dos espectros de RMN foi evidenciado a presença de ácido carboxílico e dicarboxílico, caracterizando o composto formado como ácido succínio. A análise HPLC-DAD em um comprimento de onda em λ_{\max} 365 nm mostrou

apenas um pico com um tempo de retenção (RT) em 15,7 min. Este pico mostrou absorções características em espectros de UV dos compostos flavonoides.

Embora existam muitos estudos científicos utilizando a espectroscopia NIR utilizando técnicas de classificação, poucos são os trabalhos que fizeram uso dessas técnicas para estudos com *Agave* spp. Apenas dois artigos foram encontrados, mas um deles não foi possível obter resultados satisfatórios, tais artigos estão detalhados a seguir.

RIJAL *et al.*, (2016) desenvolverem um método para avaliar a qualidade da folha de *Agave tequilana* para produção de bioetanol utilizando espectroscopia NIR com transformada de Fourier. Os espectros NIR foram coletados a partir de folhas de diferentes estágios de maturidade, em diferentes estações do ano e de dois locais diferentes. A região espectral escolhida foi de 1.100 a 2.500 nm, com uma resolução de 12.500.000 e com 64 varreduras por amostra. A técnica quimiométrica utilizada foi a regressão por mínimos quadrados parciais (PLSR), as estatísticas de previsão foram $r^2 = 0,53-0,66$ e RMSEP = 2,63-2,18% para teor de matéria seca da folha fresca e $r^2 = 0,33-0,56$ e RMSEP = 1,88-2,45% para teor de açúcar solúvel total.

Mellado-Mojica e López (2015) desenvolveram um método para identificação, classificação por espectroscopia NIR para avaliar xaropes de *Agave* spp. a partir de amostras de 43 adoçantes naturais. Foram feitas 32 varreduras com resolução de 2.500.000 nm no modo de transmitância (%T) para cada espectro na faixa de 1.000 a 2.500 nm. Não houve diferenças significativas entre os adoçantes naturais na região de 1.250-2.500 nm. A região de carboidratos NIR (1.923-2.381) não foi capaz de distinguir entre os xaropes, apesar das diferenças de origem (dados não mostrados). Foi realizada uma PCA dos edulcorantes naturais com base nos espectros da região de carboidratos NIR; no entanto, não foi possível desenvolver modelos para agrupá-los, seja entre os xaropes naturais de diferentes origens ou entre os xaropes das duas espécies de *Agave* spp. (*A. tequilana* e *A. salmiana*).

2.5. Espectroscopia no Infravermelho

A crescente demanda por melhoria da qualidade e monitoramento dos produtos pelas indústrias, em estudos acadêmicos e de pesquisa aplicada, tem culminado na substituição de técnicas analíticas demoradas, de custos operacionais elevados e dispendiosas. A Espectroscopia no Infravermelho é uma técnica que tem se mostrado

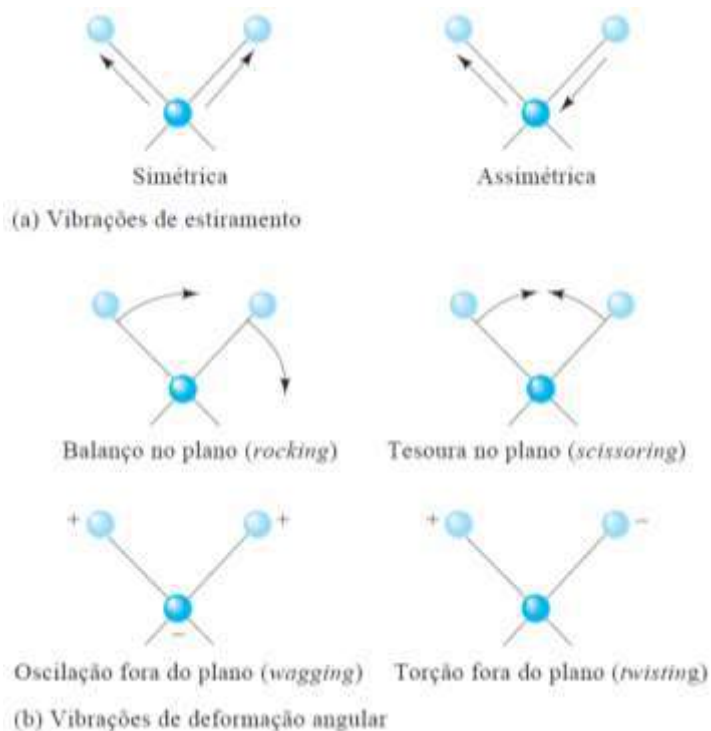
versátil, sua rapidez permite medições mais frequentes de amostras sólidas ou líquidas com um mínimo de preparo de amostras, permitindo em algumas situações aplicações não destrutivas e não invasivas. Essas características a tornam uma técnica atraente para a caracterização direta ou indireta de produtos naturais e sintéticos (PASQUINI, 2003; SIESLER, 2008; XIAOBO *et al.*, 2010).

A região do infravermelho no espectro eletromagnético compreende a faixa de comprimento de onda entre 780 nm a 1.000.000 nm, este sendo subdivida em três regiões: o infravermelho próximo – NIR (780 nm a 2.500 nm), o infravermelho médio – MIR (2.500 nm a 50.000 nm) e o infravermelho distante – FAR (50.000 nm a 1.000.000 nm) (SKOOG, HOLLER, NIEMAN, 2020).

É necessário que uma molécula possua diferença no momento de dipolo para que ocorra absorção da radiação infravermelha durante os movimentos rotacionais e vibracionais. Na molécula de ácido clorídrico, como há diferença na distribuição de cargas, logo, a molécula apresenta momento dipolar significativo. O que não ocorre em moléculas diatômicas homonucleares, por exemplo, o O₂ não absorve radiação infravermelha, pois não possui variação no momento de dipolo (SKOOG, HOLLER, NIEMAN, 2020).

Os átomos nas moléculas não estão fixos, devido a variação relacionados a inúmeros tipos de vibrações e rotações que ocorrem em torno da molécula. Essas vibrações são classificadas em estiramento e deformação. As vibrações de estiramento envolvem movimentos de átomo na direção da ligação, enquanto as vibrações de deformação envolvem movimentos que resultam em mudanças correspondentes no ângulo de ligação (SKOOG, HOLLER, NIEMAN, 2020; THOMPSON, 2018). Na **Figura 3** visualiza-se essas vibrações (SKOOG *et al.*, 2009).

Figura 3 - Diferentes tipos de modo vibracionais para uma molécula triatômica heteronuclear **(a)** - Vibrações de estiramento, **(b)** - Vibrações de deformação angular.



Fonte: (SKOOG *et al.*, 2009.)

Um tratamento da mecânica quântica mostra que a energia vibracional pode ter apenas certos valores discretos chamados níveis de energia. Para um oscilador harmônico, esses níveis de energia são dados de acordo com a **Equação 1** (SIESLER, 2008).

$$E_v = h\nu \left(1 + \frac{1}{2}\right) \quad 1$$

Em que:

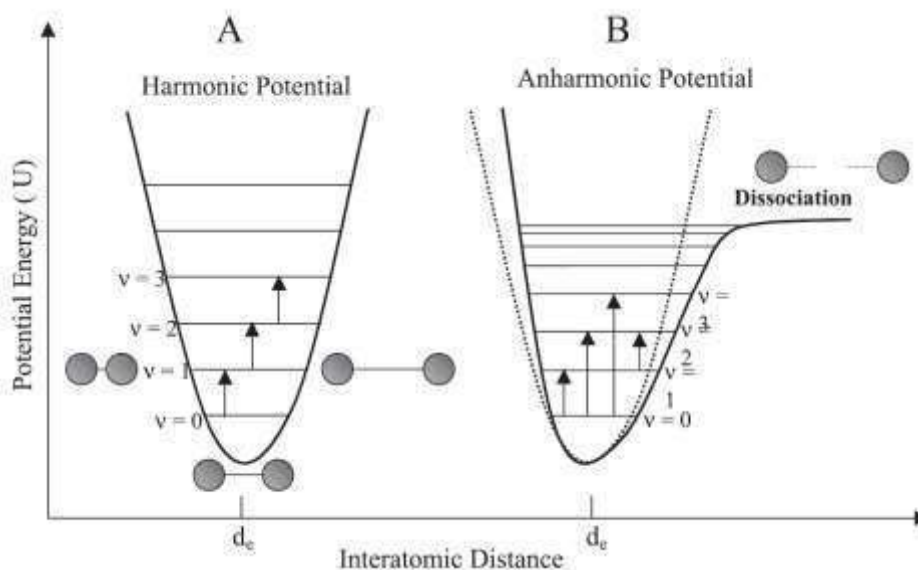
E_v = energia vibracional;

h = constante de Planck;

ν = frequência vibracional clássica.

No oscilador harmônico mecânico quântico, uma restrição adicional diz respeito ao número quântico vibracional, que só pode mudar em uma unidade, $\Delta v = \pm 1$. A maioria das moléculas à temperatura ambiente existe no nível vibracional $v = 0$ e conseqüentemente a transição permitida é $v = 0 \rightarrow v = 1$, chamada de transição fundamental, conforme a **Figura 4a** (PASQUINI, 2003; SIESLER, 2008).

Figura 4 - (a) - Diagrama de energia potencial para os osciladores harmônicos, **(b)** – Anarmônico.



Fonte: (PASQUINI, 2003).

Observações experimentais evidenciam que as moléculas não são osciladores ideais, mas sim osciladores anarmônicos (**Figura 4b**), sendo responsável pelo aparecimento de sobretons que diferem em duas ou três unidades de número quântico vibracional ($\Delta v = \pm 2, \pm 3 \dots$) e bandas de combinação de níveis vibracionais (SIESLER, 2008; SIMÕES, 2008).

2.5.1. Espectroscopia no Infravermelho Próximo (NIR)

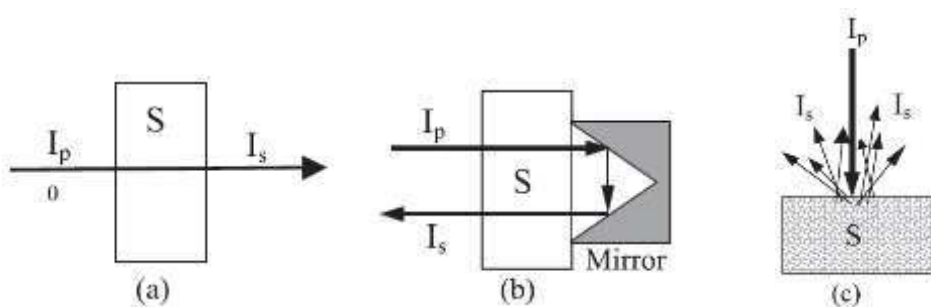
A Espectroscopia no infravermelho próximo é uma técnica analítica que compreende a faixa de absorção de 780 a 2500 nm do espectro eletromagnético, essa interação da radiação de grupos funcionais (-CH, -OH, -NH e -SH) com a matéria está relacionada a excitação de sobretons e combinações de modos vibracionais com a faixa espectral NIR (PASQUINI, 2018).

As faixas espectrais de absorção dos espectros NIR são de 10 a 100 vezes menores que o espectro infravermelho médio, tornando esta região única e marcadamente diferente das outras regiões. Isso confere uma maior profundidade de penetração da radiação. Assim, temos a possibilidade de realizar análises diretas de absorventes fortes e amostras com espalhamento, como líquidos turvos ou sólidos (BEĆ, HUCK, 2019; PASQUINI, 2018; SIMÕES, 2008; OZAKI *et al.*, 2021).

Na **Figura 5** podemos ver os modos de transmitância, transfectância e reflectância difusa que são utilizados como medição em espectroscopia NIR. A transmitância (**Figura 5a**) é o modo comumente utilizado na espectroscopia UV-VIS

convencional, com caminho óptico variando de 1 a 50 mm. A transfectância (**Figura 5b**) é frequentemente usada quando são empregadas sondas de feixe óptico, sua diferença para a transmitância está em dobrar o caminho óptico, devido à passagem do feixe duas vezes pela amostra. O modo de reflectância difusa (**Figura 5c**) a dispersão e a absorção por grânulos sólidos contribuem para alterar a intensidade do sinal (JAMRÓGIEWICZ, 2012; PASQUINI, 2003; PASQUINI, 2018).

Figura 5 - Modos de medidas na região NIR. **(a)** transmitância, **(b)** transfectância e **(c)** refletância difusa.



Fonte: (PASQUINI, 2003).

Um dos desafios para a espectroscopia NIR, é obter apenas informações úteis dos grandes conjuntos de dados e da complexidade dos espectros, além da interpretação do fenômeno com as propriedades físicas e químicas da espécie de interesse, o que prejudica a interpretação direta dos dados. Portanto, a utilização de ferramentas matemáticas e estatísticas, a partir de técnicas quimiométricas para evidenciar a resposta espectral obtida são indispensáveis (MANLEY, 2014; OSBORNE, 2006; JAMRÓGIEWICZ, 2012; SIMÕES, 2008).

2.6. Quimiometria

Os primeiros trabalhos sobre Quimiometria começaram a ser publicados no início da década de 70, junto ao desenvolvimento da computação científica e envolveu principalmente o uso de métodos estatísticos multivariados para análise de dados analíticos (BOVENS, M. *et al.* 2019). O termo “quimiometria” incorpora uma vasta gama de métodos matemáticos e algoritmos que podem ser aplicados a determinados conjuntos de dados. Esses métodos podem resultar na formulação de modelos

qualitativos e quantitativos, muitas vezes podem identificar relações subjacentes existentes entre variáveis (CHAPMAN *et al.*, 2020; DUPONT *et al.*, 2020).

Entretanto, a Química Analítica e as demais áreas dentro da Química, não são as únicas disciplinas a se beneficiarem das técnicas quimiométricas. Uma vasta gama de trabalhos publicados em outras áreas como ciência forense (SAUZIER *et al.*, 2021), análise ambiental (CHAPMAN *et al.*, 2020), indústria farmacêutica (TEIXEIRA, 2017), no controle de qualidade (MOREIRA, 2018), na área médica e alimentar (FERREIRA, M., 2015), revelam sua natureza multidisciplinar. Por existir muitos métodos quimiométricos, cada método deve ser utilizado de acordo com o tipo de estudo: caracterização, discriminação, desenvolvimento de modelo quantitativos, dentre outros (KUMAR, SHARMA, 2018).

Apesar de possuir uma grande variedade de aplicações, as técnicas quimiométricas são divididas em três categorias. Uma delas é a técnica de reconhecimento de padrões. Pode ser subdividida em duas categorias como os métodos não supervisionados, utilizada para reconhecer as tendências dos conjuntos de dados (análise exploratória de dados), e os métodos supervisionados que tem por objetivo utilizar essas tendências criando modelos para classificação e predição de conjuntos desconhecidos. Os métodos de regressão (calibração multivariada) são utilizados para prever quantitativamente as propriedades de interesse em uma amostra. E por fim, as técnicas de planejamento e otimização experimental (SAUZIER *et al.*, 2021)

Desde o desenvolvimento da quimiometria até recentemente diversos tipos de ferramentas com objetivo específico (algoritmos de seleção de amostras, seleção de variáveis, dentre outros) possibilitaram o processamento e interpretação de dados que antes não havia a possibilidade de serem analisados (SIMÕES, 2008).

2.6.1. Análise de Componentes Principais (PCA)

A Análise de Componentes Principais (PCA) é provavelmente a técnica quimiométrica multivariada mais difundida e, devido à importância das medições multivariadas em química, é considerada por muitos como a técnica que mudou mais significativamente a visão química da análise de dados (LEE, JEMAIN, 2021; MISHRA *et al.*, 2017).

A Análise de Componentes Principais (PCA) possui como sua ideia central reduzir a dimensionalidade de um grande conjunto de dados originais que estão inter-relacionadas, gerando um novo conjunto de dados. De forma a manter as informações mais relevantes, facilitando a interpretação e a visualização das possíveis relações existentes entre as amostras. Entretanto, a PCA não é uma técnica utilizada para classificação de amostras (FERREIRA, M., 2015; MISHRA *et al.*, 2017; XU, *et al.*, 2020).

A equação abaixo fornece o modelo que representa a transformação da matriz de dados em um modelo bilinear:

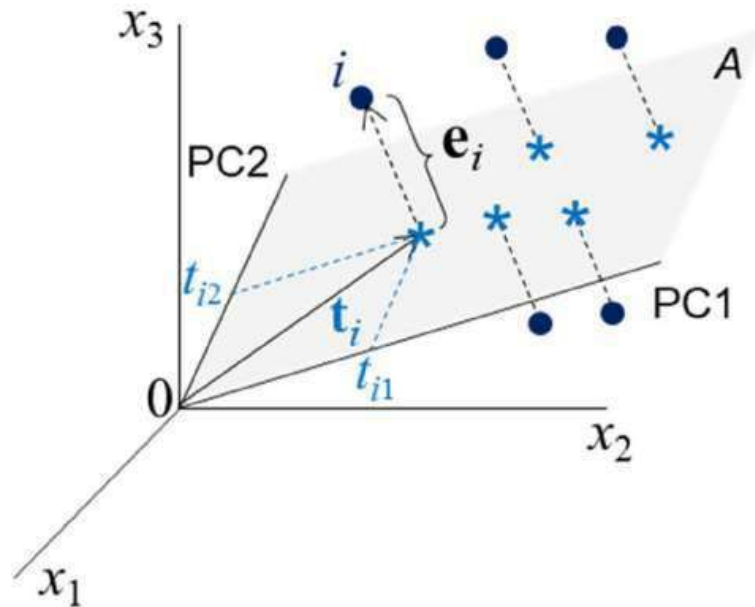
$$\mathbf{X} = \mathbf{T}\mathbf{L}^T + \mathbf{E} \quad (2)$$

No qual, \mathbf{X} representa o conjunto original dos dados, \mathbf{T} é a matriz de scores (novas coordenadas no novo sistema de eixos), \mathbf{L} a matriz dos pesos (Loadings, que são os cossenos dos ângulos entre os eixos originais e as componentes principais) e \mathbf{E} é a matriz dos erros (LEE, JEMAIN, 2021).

O novo conjunto de variáveis (componentes principais) é uma combinação linear das variáveis originais. Os novos eixos são ortogonais entre si e são organizados em ordem decrescente de acordo com a quantidade de variação que descrevem. Então podemos dizer que o PCA tenta agrupar as variáveis que são altamente correlacionadas em uma nova variável chamada componente principal (MISHRA *et al.*, 2017).

Os dados originais são projetados em um novo sistema de coordenadas no qual o primeiro eixo (PC1), está disposta ao longo do eixo na direção no qual se explica a maior variação dos dados, o segundo eixo (PC2) corresponde à direção ao longo do eixo que explica a maior variação dos dados que não foram explicadas pela PC1, assim sucessivamente, conforme pode ser visualizado na **Figura 6**. A quantidade total de PC's é igual ao número de dimensões do conjunto de dados originais. Entretanto, os últimos PC's podem ser omitidos, pois sua variância explicada são as menores (KHERIF, LATYPOVA, 2020).

Figura 6 - Representação gráfica da projeção das amostras no subespaço A gerado pelas duas componentes principais, PC1 e PC2. \mathbf{t}_i é o vetor de escores definido no plano das PC; t_{i1} e t_{i2} são as coordenadas (os escores) de i nesse plano e \mathbf{e}_i é o vetor de resíduos, ortogonal ao plano A.



Fonte: (FERREIRA, M., 2022).

A PCA tem muitas propriedades atraentes como: velocidade computacional, pode-se obter uma melhor visão dos dados por meio das imagens geradas, evita problemas de multicolinearidade e reduz o risco de superajuste, além de ser uma técnica utilizada para compressão de imagens (KHERIF, LATYPOVA, 2020).

A PCA tem como seus principais objetivos o de verificar as possíveis relações entre objetos e agrupá-los de acordo com suas semelhanças, o que dá a possibilidade de identificar amostras anômalas, que não fazem parte de nenhuma das categorias conhecidas. Extrair as informações mais importantes dos conjuntos de dados, reduzindo, assim, a dimensão dos dados sem perda significativa das informações (MISHRA *et al.*, 2017). Contudo, na PCA os resultados de saída requerem mais esforço, devido a transformação matemática dos dados, como alternativa, o algoritmo de colônia de formiga acoplado a análise discriminante linear (ACO-LDA) apresenta maior facilidade de interpretação dos resultados.

2.6.2. Algoritmo de Colônia de Formiga acoplado a Análise Discriminante Linear (ACO-LDA)

A Análise Discriminante Linear (LDA) (do inglês, *Linear Discriminant Analysis*) foi proposto por Fisher (1936), com a finalidade de resolver problemas relacionadas a

duas classes-alvo, a partir de hiperplanos como fronteiras de decisão lineares. Portanto, o objetivo é projetar as amostras em um espaço onde as amostras da mesma classe fiquem o mais próximas possível, enquanto as amostras de classes diferentes fiquem o mais distantes possível. Para isso, tanto a dispersão entre as classes quanto a dispersão dentro de cada classe são levadas em consideração durante o processo de projeção (PONTES *et al.*, 2020; ZHU *et al.*, 2022).

Apesar do surgimento de dados com alta dimensionalidade, o LDA permanece como uma das abordagens mais populares para enfrentar uma variedade de problemas de classificação (SAFO, AHN, 2016). Entretanto, o LDA apresenta algumas desvantagens tais como: dificuldade quando o número de classes se torna grande (maior que 5), é sensível a colinearidade, só pode ser aplicado quando o número de amostras de treinamento for maior do que o número de variáveis consideradas no modelo. No entanto, quando lidamos com conjuntos de dados com alta dimensionalidade, é crucial adotar procedimentos adequados de seleção de variáveis (FERREIRA, M., 2015; MILANEZ, PONTES, 2014).

A utilização do algoritmo de Otimização por Colônias de Formigas (ACO), (do inglês, *ant colony optimization*) acoplado ao LDA foi apresentado pela primeira vez por Pontes *et al.* (2020), o ACO-LDA utiliza uma estratégia de busca não específica, inspirada no comportamento cooperativo das formigas. Nesse algoritmo, os feromônios são depositados aleatoriamente nos elementos do vetor que representam as variáveis mais relevantes. Essa imitação do comportamento cooperativo aumenta a chance de descartar variáveis que não são informativas, o que leva à construção de modelos mais parcimoniosos.

No algoritmo ACO-LDA, são utilizadas formigas binárias padronizadas, cujos comprimentos correspondem ao número de variáveis presentes na matriz de resposta instrumental \mathbf{X} ($n \times j$), onde n representa o número de amostras e j representa o número de variáveis. Essas formigas são codificadas de forma binária, adotando os valores "1" ou "0" para indicar, respectivamente, se uma variável foi selecionada ou não selecionada (PONTES *et al.*, 2020).

A construção do modelo ocorre a partir de cada formiga do formigueiro de inicial, como descrito pela **Equação 3**:

$$\mathbf{G}_{custo} = \frac{1}{dof} \sum_{K=1}^{K_{treinamento}} \mathbf{g}_K \quad 3$$

onde G_{custo} representa o risco médio de classificação pelo subconjunto de variáveis, o dof é o valor o grau de liberdade, representado por:

$$dof = K_{treinamento} - l - tc \quad 4$$

onde $K_{treinamento}$ é o número de amostras no conjunto de treinamento, l é o número de formigas incluídas para cada formiga e tc é o número de amostras no conjunto de dados externo. O g_K na **Equação 3** é o risco de classificação incorreta do conjunto de treinamento (PONTES *et al.*, 2020).

2.6.3. Calibração Multivariada

Nos métodos de calibração multivariada, duas ou mais respostas do instrumento são relacionadas com a propriedade de interesse, isso nos possibilita análises de amostras que apresentem interferentes, desde que, a amostra de calibração também possua tais interferentes (OLIVIERI, 2018).

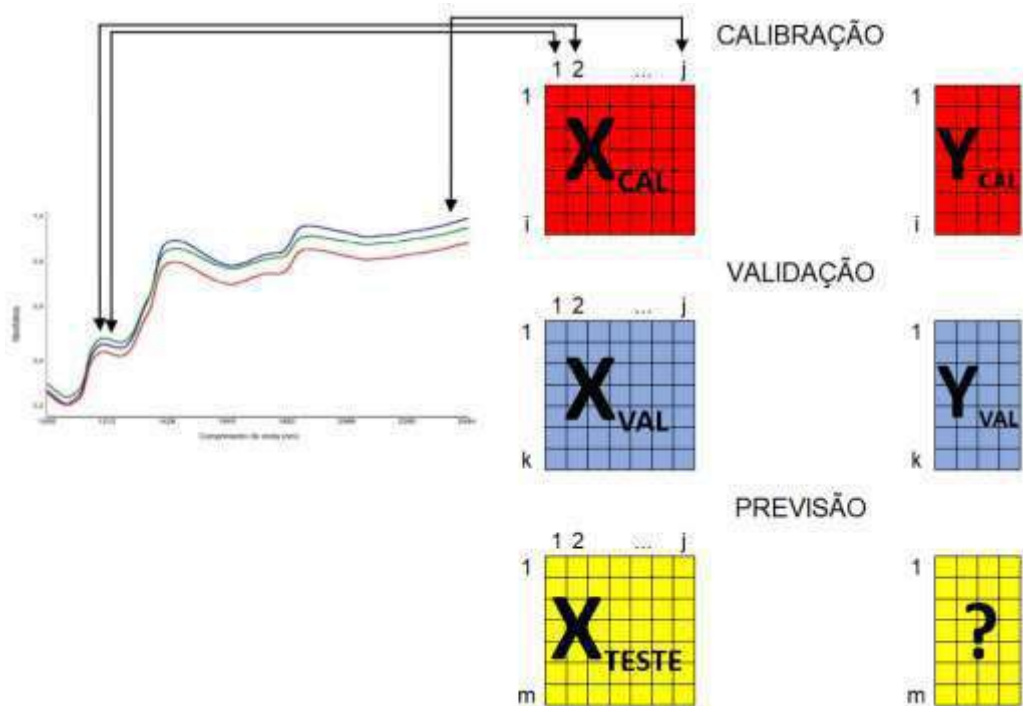
O processo de calibração é formado por três fases: a calibração, a validação e a previsão. Os modelos matemáticos são construídos na fase de calibração. O modelo de calibração é construído a partir de um conjunto de amostras experimentais no qual as concentrações e espectros do analito são conhecidos. A ideia central para essa construção desse conjunto de calibração é que o modelo aprende a relação entre o espectro e a concentração (SIMÕES, 2008; OLIVIERI, 2018).

Após construído o modelo de calibração, este deve ser testado frente a um conjunto de amostras de validação independente, no qual as concentrações do analito que estão presentes sejam diferentes daqueles empregados na calibração (conhecido como validação externa do modelo e o conjunto de amostras é denominado conjunto de teste), com a finalidade de testar a capacidade preditiva do modelo, comparando-se as concentrações estimadas do analito para o conjunto de validação por meio de ferramentas estatísticas (OLIVIERI, 2018; FERREIRA, M., 2015):

Na fase de previsão deve-se se atentar em alguns cuidados tanto no preparo das amostras, quanto na coleta dos dados e instrumentos utilizados, para que sejam obtidos de maneira semelhante ao do conjunto de calibração. Portanto, os modelos matemáticos criados e validados são aplicados com o objetivo de prever as

concentrações ou alguma propriedade de amostras desconhecidas (**Figura 7**) (SIMÕES, 2008; FERREIRA, M., 2015).

Figura 7 - Organização dos dados da calibração multivariada.



Fonte: Adaptado de (SIMÕES, S. da S., 2008).

Espectros de “ l ” amostras do conjunto de calibração e de “ k ” amostras de validação são obtidas em “ j ” valores de comprimento de onda diferentes, formando uma matriz X_{cal} ($i \times j$) para a fase de calibração e outra matriz X_{val} ($k \times j$) para a fase de validação. Duas matrizes Y_{cal} ($i \times q$) e Y_{val} ($k \times q$) também são formadas com “ q ” colunas, relacionadas aos diferentes componentes presentes nas amostras e cujos valores de concentração são conhecidos geralmente por métodos de análise de referência (SIMÕES, 2008).

Alguns métodos de regressão vêm sendo utilizados para a construção de modelos de calibração multivariada, um deles é a Regressão por Mínimos Quadrados Parciais (PLS). É uma técnica robusta e apropriada quando a resposta (variáveis dependentes) exibem colinearidade com muitas variáveis preditoras (variáveis independentes) (GEBREMICAEL *et al*, 2019).

2.6.3.1. Regressão por Mínimos Quadrados Parciais (PLS)

A Regressão por Mínimos Quadrados Parciais (PLS) oferece possibilidades de interpretação que podem ser melhor explicadas e ilustradas (AKARACHANTACHOTE, *et al.*, 2014). O PLS é uma técnica de regressão multivariada robusta que é apropriada quando a resposta (variáveis dependentes) exibe colinearidade com muitas variáveis preditoras (independentes) (GEBREMICAEL, *et al.*, 2019). Como as novas variáveis não são mais componentes principais, elas são normalmente chamadas de variáveis latentes (LVs) ou componentes PLS. Isso resulta em uma decomposição bilinear de ambas as matrizes centradas X e Y em matrizes de escores e cargas análoga à decomposição obtida em PCA (SAEYS, *et al.*, 2019).

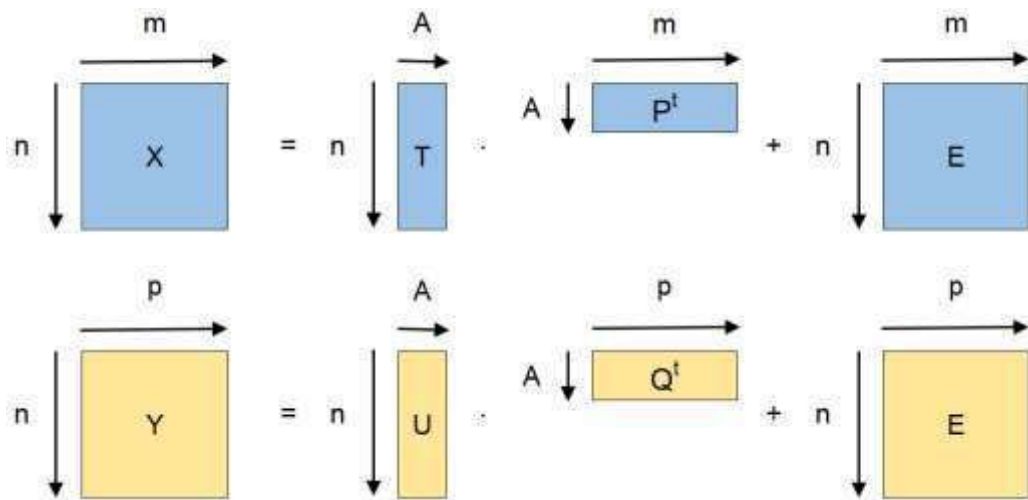
O modelo PLS é construído em uma única etapa, na qual as informações da matriz X e da propriedade de interesse são consideradas durante a decomposição e a compressão dos dados (**Equações 5 e 6**). Cada variável latente, que relaciona X e a propriedade de interesse y , é obtida maximizando a covariância entre escores t , da matriz X , e y . Com essa definição dos fatores, estamos otimizando a decomposição na direção da previsão da propriedade y , a decomposição das matrizes X e Y são observadas na **Figura 8**. O algoritmo NIPALS e sua versão não ortogonal, foram os primeiros algoritmos PLS usados para construir modelos de regressão (FERREIRA, M., 2015; SANTOS, 2013).

$$X = TP^t + E \quad (5)$$

$$Y = UQ^t + F \quad (6)$$

Em que: T e U são as matrizes dos escores e P e Q são as matrizes dos pesos das matrizes X e Y . E é a matriz de resíduos espectrais e F é a matriz dos resíduos de concentração.

Figura 8 - Decomposição das matrizes X e Y no algoritmo do PLS.



Fonte: Adaptado de (FERREIRA, M., 2022).

Por fim, o modelo PLS consiste em relacionar linearmente os escores da matriz **X** com os escores da matriz **Y** (SIMÕES, 2008), como indica as **Equações 7 e 8**.

$$\mathbf{U} = \mathbf{B}\mathbf{T}^t + \mathbf{G} \quad (7)$$

$$\mathbf{Y} = \mathbf{B}\mathbf{T}\mathbf{Q}^t + \mathbf{H} \quad (8)$$

Em que: **B** é a matriz dos coeficientes de regressão; **G** é a matriz de resíduos dos escores e **H** é a matriz de resíduos de concentração.

Uma das etapas cruciais envolve escolher adequadamente o número de variáveis latentes. Essa escolha tem um impacto significativo na qualidade e desempenho do modelo resultante. Se for escolhido um número insuficiente de variáveis latentes, pode-se enfrentar um problema de subajuste, o que significa que o modelo não conseguirá aproveitar completamente todas as informações úteis disponíveis para sua construção, nesse caso, estaria deixando-se de considerar aspectos relevantes dos dados. Por outro lado, se escolhe-se um número excessivo de variáveis latentes, correse o risco de enfrentar um problema de sobreajuste. Isso ocorre quando o modelo incorpora informações que não estão diretamente relacionadas à propriedade de interesse que almeja modelar. Essas informações adicionais podem incluir ruídos espectrais ou outras fontes de variação não relevantes para o modelo (SANTANA *et al.*, 2020). A seleção do número de variáveis latentes (LV) é uma etapa crucial. Essa seleção é realizada com base no desempenho do

modelo, como o erro médio quadrático de validação (RMSEV) ou o erro médio quadrático de validação cruzada (RMSECV) (DINIZ, 2020).

Para a construção dos modelos é necessário dividir os conjuntos das amostras em conjuntos de calibração, predição e validação. Para essa divisão uma alternativa é utilizar o algoritmo SPXY para particionar o conjunto de amostras com base nas distâncias X-y conjuntas (DINIZ, 2020; TIAN *et al.*, 2018). Quando o número de amostras disponíveis é limitado, é recomendado realizar uma divisão das mesmas em conjuntos de calibração e predição, seguida pela aplicação de validação cruzada completa (CV, do inglês cross-validation) nas amostras de calibração. Pois permite avaliar a precisão e a capacidade preditiva dos modelos que foram criados (DINIZ, 2020; WULANDARI; PERMANA; KRISTININGRUM, 2020).

2.6.4. Regressão por Mínimos Quadrados Parciais em intervalos (iPLS)

O iPLS é um método que utiliza a divisão do espectro completo em subintervalos equidistantes para construir modelos PLS por meio da validação cruzada. Cada subintervalo é avaliado individualmente e também é calculado um RMSECV médio para o modelo completo do espectro. A região com o menor RMSECV é selecionada como a mais relevante (CHEN *et al.*, 2018).

Os modelos construídos com base nos diferentes subintervalos geralmente possuem um número distinto de variáveis latentes em comparação com o modelo completo, permitindo capturar a variação relevante específica de cada região. O iPLS fornece uma visão geral das informações úteis presentes nos subintervalos, concentrando-se nas regiões espectrais importantes e evitando interferências de outras regiões. Além disso, permite a comparação entre modelos locais e o modelo global. Essa abordagem fornece uma compreensão abrangente dos dados e é útil na interpretação dos resultados (CHEN *et al.*, 2018; GOMES, 2012).

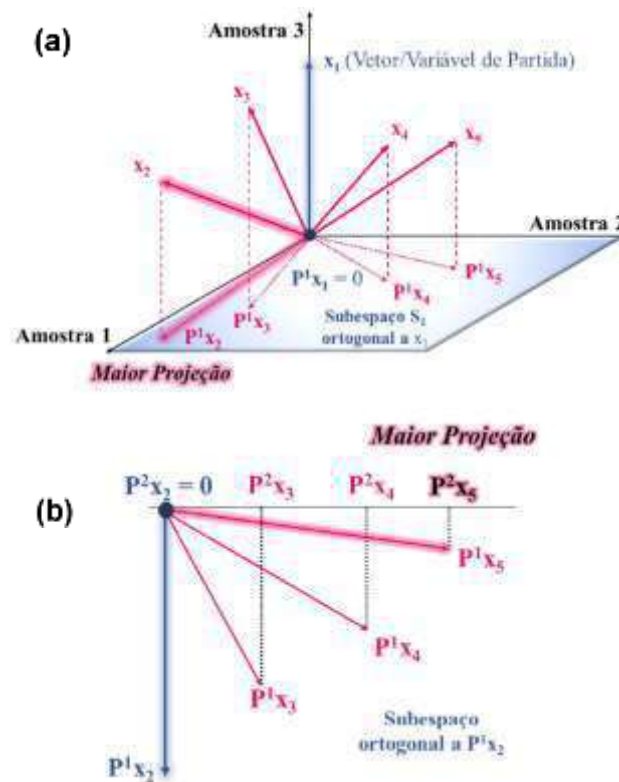
2.6.5. Algoritmo das Projeções Sucessivas em intervalos para Regressão PLS (iSPA-PLS)

O Algoritmo das Projeções Sucessivas (do inglês, *Successive Projection Algorithm* (SPA)) foi proposto em 2001, por Araújo *et al.* (2001), sendo aplicado de forma altamente eficaz para resolver problemas de multicolinearidade em modelos de

regressão linear múltipla (do inglês, *Multiple Linear Regression* (MLR)). Basicamente, o SPA envolve a execução de operações de projeção na matriz de calibração X_{cal} ($K_c \times J$), na qual as linhas representam K_c amostras de calibração e as colunas representam J variáveis espectrais (MILANEZ, 2017). A **Figura 9** fornece o procedimento de projeções sucessivas das variáveis até a obtenção de uma cadeia com o melhor subconjunto.

A partir da coluna inicial x_1 , o SPA determina, no subespaço S_0 ortogonal a x_1 , qual das colunas restantes possui a maior projeção. Essa coluna, chamada de x_2 , pode ser considerada como aquela que contém a maior quantidade de informação não incluída em x_1 . Na próxima iteração, o SPA restringe a análise ao subespaço S_0 , usando x_2 como a nova coluna de referência, e continua seguindo os passos descritos. Dessa forma, o critério de seleção do SPA favorece a redução da colinearidade entre as variáveis. Além disso, o SPA permite selecionar no máximo K_c variáveis, uma vez que a cada operação de projeção, um grau de liberdade é removido da matriz X (HONORATO, 2006).

Figura 9 - Sequência de projeções realizadas pelo SPA com $K_c = 3$ e $j = 5$. Resultado da primeira (a) e segunda (b). Neste exemplo, o conjunto de variáveis selecionadas inicia com x_1 e deverá ser $\{x_1, x_2, x_5\}$.



Fonte: Milanez (2017).

O melhor subconjunto é escolhido de acordo com a raiz quadrada do erro médio para o conjunto de validação (RMSEV), como descrito na **Equação 13**.

$$RMSEV = \sqrt{\frac{1}{K_{val}} \sum_{k=1}^{K_{val}} (y_V^k - \hat{y}_V^k)^2} \quad (13)$$

onde y_V^k e \hat{y}_V^k correspondem, respectivamente, ao valor de referência e valor previsto, para o parâmetro de interesse na k-ésima amostra da validação K_{val} é o número de amostras de validação (SOARES *et al.*, 2013).

3. MATERIAIS E MÉTODOS

3.1. Local dos ensaios

A parte experimental de presente trabalho foi desenvolvida no Laboratório Avançado de Tecnologia Química (LATECQ) da Embrapa Algodão, em Campina Grande, PB. Já a construção dos modelos foi realizada no laboratório de Síntese Inorgânica e Quimiometria (labSIQ) da Universidade Estadual da Paraíba, Campus I, Campina Grande - PB.

3.2. Coleta de folhas de *Agave* spp. e extração da fração líquida

A coleta das amostras foi realizada no Banco de Germoplasma de *Agave* da EMBRAPA com plantas em idade de 51 meses de cultivada na Estação Experimental da EMBRAPA em Monteiro-PB. As folhas de *Agave* spp. coletas por pessoal especializado foram encaminhadas laboratório para processamento, caracterizações físico-químicas e instrumentais. A massa das folhas foi medida e, depois tiveram seus espinhos de ponteiro removidos. Posteriormente, as folhas foram lavadas com água deionizada e cortadas com auxílio de uma faca de aço inoxidável, em cubos de tamanhos de cerca de 2,00 cm³. As amostras com tamanho reduzido foram armazenadas em sacos plásticos de 10,00 kg dentro de um freezer, sendo retiradas quantidades, conforme a necessidade para descongelamento para os ensaios experimentais.

Foram coletadas folhas de 6 diferentes genótipos de *Agave* denominados de acordo com a **Tabela 1**. Cada genótipo foi constituído por sete plantas (repetições), exceto para o genótipo D do qual foi composto de cinco plantas. Na **Tabela 1** também pode-se identificar a codificação utilizada para a construção dos modelos quimiométricos.

Os extratos das amostras das folhas dos seis genótipos de *Agave* spp. foram extraídos a partir de três folhas do terço médio de cada planta. Cada folha foi selecionada observando a ausência de doenças ou ataques de insetos.

Tabela 1 - Amostras de genótipos de *Agave* spp. do Banco de Germoplasma de *Agave* da EMBRAPA com plantas em idade de 51 meses de cultivada na Estação Experimental da EMBRAPA em Monteiro-PB utilizadas na análise.

Genótipos	Codificação para modelos	Número de amostras
A	1	7
B	3	7
C	7	7
D	10	5
E	14	7
F	15	7

Fonte: Elaborado pelo autor, 2023

3.3. Caracterização físico-química dos extratos processados de *Agave* spp.

3.3.1. Preparo das amostras para a análise de degradação

As amostras foram removidas do freezer e descongeladas para em seguida serem trituradas em liquidificador de uso doméstico. O material triturado foi filtrado de forma manual utilizando um tecido 100% de algodão previamente esterilizado. Um volume de 50,0 mL do filtrado foi transferido para garrafas PET de 100mL contendo conservante, com os seguintes tratamentos utilizando um conservante químico nas seguintes concentrações: 0%, 0,01%, 0,05%, 0,1% e 0,5% (m/v). Os extratos coletados de todas as plantas individuais seguiram o mesmo procedimento. Informações detalhadas do conservante utilizado não poderão ser fornecidas pois o mesmo está sob proteção de patente.

3.3.2. Determinação da perda de massa dos extratos líquidos

A determinação da massa de cada extrato produzido foi realizada em balança analítica da marca CELTAC JA3003N. A massa da amostra foi determinada diariamente até o 21º dia e depois no 33º, 43º e 51º dias. Antes da pesagem o recipiente foi aberto para a saída dos gases gerados pela fermentação espontânea.

3.3.3. Determinação de condutividade elétrica em extratos líquidos

As medidas para determinação de condutividade foram realizadas diariamente durante 51 dias utilizando o equipamento da marca MS TECNOPON INSTRUMENTAÇÃO e modelo mCA150. A calibração foi realizada com uma solução padrão de condutividade de 146,9 $\mu\text{S}/\text{cm}$. Nas medidas, a cada mudança de amostra o eletrodo foi lavado abundantemente com água deionizada e seco com papel lenço macio.

3.3.4. Determinação de pH em extratos líquidos

Na determinação de pH utilizou-se um instrumento com eletrodo seletivo para íons H^+ (marca JKI, modelo JK-PHM-005) com ajuste de calibração com soluções tampão de pH 4,0 e 7,0. As medidas foram realizadas diariamente de forma análoga ao item da seção 3.3.2. O eletrodo de íon seletivo foi inserido parcialmente nos frascos dos extratos e a leitura registrada. Entre as medidas, o eletrodo foi lavado com água deionizada e seco com papel lenço macio, sem deixar resíduos.

3.3.5. Determinação de acidez total

A determinação da acidez total foi realizada por titulação com soluções padronizadas 0,10 molL^{-1} de Hidróxido de Sódio PA (ACS Científico) e 1,0% (m/v) de solução alcoólica de Fenolftaleína PA (Proquimios) como solução indicadora. A solução de NaOH foi padronizada com Biftalato de Potássio PA (Proquimios).

3.3.6. Determinação do teor de água

A determinação do teor de água foi realizada a partir da medida da massa de 10,00 g da amostra de *Agave* spp. de cada planta em cápsula de alumínio com 100,00 mL de capacidade. As massas das cápsulas foram previamente medidas, depois submetidas a estufa de circulação forçada de ar, a uma temperatura de 105°C durante 2 horas (Marca Solab e Modelo SL 102). A secagem das amostras foi executada por 24 horas nas cápsulas com massas previamente aferidas.

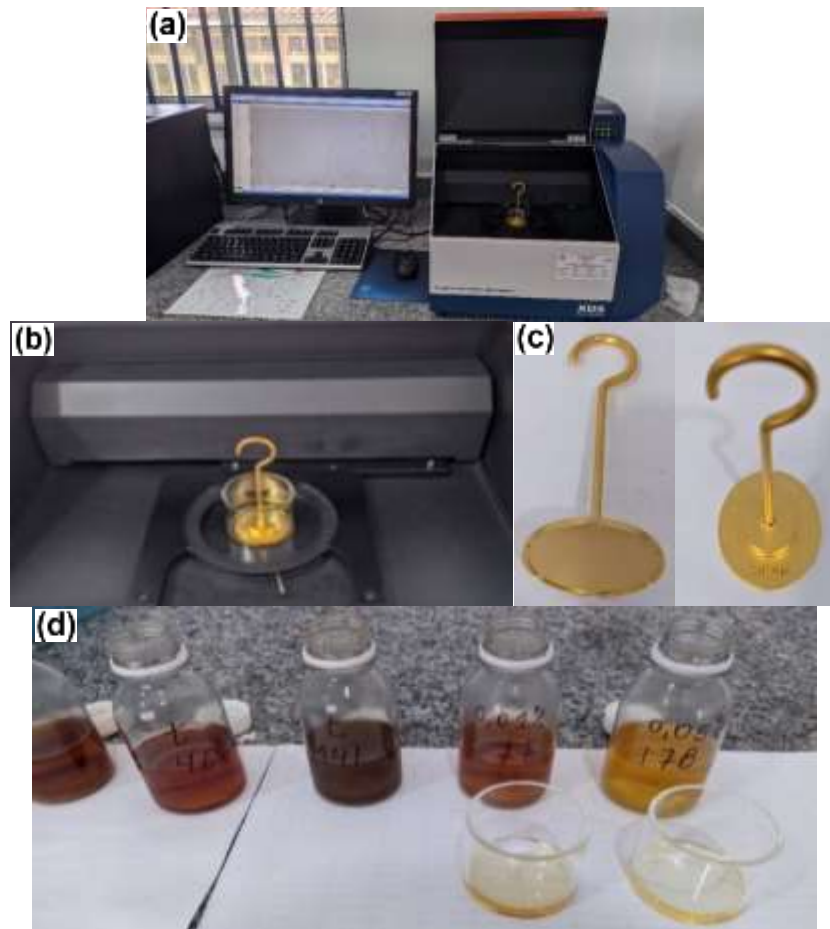
3.3.7. Determinação de sólidos totais no extrato líquido pela técnica gravimétrica

As medidas foram realizadas com cápsulas de alumínio e suas massas aferidas previamente. Em seguida, um volume de 10,00 mL de cada amostra dos extratos líquidos foi transferido para cada cápsula de alumínio, e submetidas a evaporação em banho-termostatizado (Marca Nova técnica e Modelo NT 248), até a redução total do volume. O material residual foi submetido a temperatura de 105°C em estufa de circulação forçada de ar por 2 horas (Marca Solab e Modelo SL 102). A massa final da cápsula com a amostra seca foi medida após equilíbrio térmico em dessecador contendo sílica gel e massa constante.

3.3.8. Medidas na região do Infravermelho Próximo - NIR

As medidas na região do infravermelho próximo foram realizadas nas amostras após 51 dias no estudo de degradação dos extratos dos genótipos de *Agave* spp. Foram também registrados espectros VIS-NIR dos extratos recém processados de cada genótipo. As medidas foram realizadas utilizando um instrumento VIS-NIR XDS RapidContent™ Analyzer (FossAnalytical, Hogans, Suécia), como o observado na **Figura 10a**. O intervalo de 0,5 nm, a faixa espectral de 400 a 2500 nm e 36 varreduras foram ajustados para as medidas por meio de uma célula cilíndrica de transfectância para líquidos com caminho óptico de 0,5 mm (**Figura 10c**).

Figura 10 - Sistema de medidas espectrais NIR para medidas e análise da estabilidade/degradação do extrato de *Agave* spp. **(a)** - Espectrômetro NIR. **(b)** - Cubeta de quartzo. **(c)** - Caminho óptico de 50 mm. **(d)** - Amostras do extrato de *Agave* spp.



Fonte: Elaborado pelo autor, 2023.

3.4. Desenvolvimento dos Modelos Quimiométricos

Neste trabalho são utilizadas três abordagens para o desenvolvimento dos métodos baseados na espectroscopia VIS-NIR e métodos quimiométricos. A primeira, exploratória, utilizando-se análise em componentes principais. Na qual buscou-se verificar a tendência de separação entre as 6 classes de genótipos coletados. A segunda, de reconhecimento de padrões (RP) supervisionados, na qual objetivou-se desenvolver modelos de análise discriminante linear (LDA) construídos a partir das variáveis selecionadas pelo algoritmo bioinspirado. E a terceira, com enfoque na predição da acidez do extrato do *Agave* spp., utilizando modelos de regressão por mínimos quadrados parciais (PLS) construídos com toda faixa espectral, e regiões selecionadas por algoritmos de seleção de variáveis.

Antes de efetivamente iniciar o desenvolvimento dos modelos quimiométricos, foi escolhida a faixa de trabalho excluindo-se regiões com absorvância superior a 2 u.a. Em seguida, com o objetivo de eliminar ou reduzir características espectrais indesejadas e melhorar a qualidade dos dados espectrais, foram testados alguns pré-processamentos espectrais, como a variação normal padrão (SNV) e primeira derivada de Savitzky-Golay com polinômio de primeira ordem combinando janelas de 3, 7 e 15 pontos.

Para a construção dos modelos de reconhecimento de padrão supervisionado e de calibração multivariada, os conjuntos de treinamento (ou calibração) e de teste (ou predição) foi utilizado o algoritmo SP-Xy (do inglês: *Sample set Partitioning based on joint X-y distances*) (TIAN *et al.*, 2018).

3.4.1. Modelos de classificação (ACO-LDA)

Para a construção do modelo de RP supervisionado utilizando LDA, foi utilizado o algoritmo de colônia de formigas (ACO) para selecionar as variáveis espectrais mais informativas. Utilizou-se como parâmetros para o modelo 10 formigas, 20 colônias, taxa de formigas cegas 0,35 e máximo de variáveis igual a 10.

A capacidade preditiva dos modelos ACO-LDA foi avaliada com base em figuras de mérito como sensibilidade, especificidade, taxa de classificação correta e taxa de classificação incorreta.

3.4.2. Modelos de calibração (PLS, iPLS, iSPA-PLS)

Já para a construção dos modelos de calibração multivariada foram testadas a capacidade de modelos PLS construídos com toda faixa espectral de trabalho, com os intervalos espectrais mais informativos selecionados pelo PLS (iPLS) e pelo SPA (iSPA-PLS). Para o iPLS foram determinados 20 intervalos como parâmetro. Já para iSPA-PLS foram determinados 20 intervalos e o máximo de intervalos a ser selecionados igual a 19. Durante a construção dos modelos, foi realizada uma análise de outliers utilizando o gráfico de influência (do inglês, *leverage*) em relação aos resíduos padronizados, com limites de confiança de 95%. Os modelos de regressão PLS foram validados por meio de validação cruzada completa, visando otimizar o número de variáveis latentes a serem incluídas no modelo.

Após a validação ou validação cruzada dos modelos, é fundamental avaliar seu desempenho utilizando diversas métricas. As quais incluem o coeficiente de determinação (R^2_{Val} ou R^2_{CV}) e a raiz do erro quadrático médio (RMSEV ou RMSECV) para o conjunto de validação ou validação cruzada. Além disso, é essencial avaliar a capacidade preditiva dos modelos finais em um conjunto externo de dados. Nesse conjunto, métricas adicionais como o coeficiente de determinação de predição (R^2_{Pred}), o erro quadrático médio de predição (RMSEP), o erro relativo de predição (REP) e a razão de desempenho do desvio (RPD) são utilizadas. Para obter um desempenho ideal, é desejável que os valores de RMSEP e REP sejam os mais baixos possíveis, indicando uma baixa dispersão e erro nas previsões. O coeficiente de determinação (R^2) deve estar próximo de 1, evidenciando um bom ajuste do modelo aos dados. Além disso, é preferível que a RPD seja maior que 3, sugerindo uma relação favorável entre a variabilidade da resposta medida e a variabilidade estimada pelo modelo (DINIZ, 2020).

$$R^2_{CV} = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (9)$$

$$RMSECV = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (10)$$

$$R^2_{Pred} = 1 - \frac{\sum_{j=1}^m (\hat{y}_j - y_j)^2}{\sum_{j=1}^m (y_j - \bar{y}_j)^2} \quad (11)$$

$$RMSEP = \sqrt{\frac{1}{m} \sum_{j=1}^m (\hat{y}_j - y_j)^2} \quad (12)$$

$$REP = \left(\frac{RMSEP}{\frac{1}{n} \sum_{i=1}^n y_i} \right) \times 100 \quad (13)$$

$$RPD = \frac{Sd_{Pred}}{RMSEP} \quad (14)$$

no qual que n e m representando os tamanhos das amostras nos conjuntos de calibração (i) e predição (j), respectivamente. Os valores preditos pelo modelo para as amostras são denotados por \hat{y}_i e \hat{y}_j , enquanto y_i e y_j representam os valores de referência das amostras. Além disso, \bar{y}_i e \bar{y}_j indicam os valores médios de referência

das amostras, e Sd_{Pred} representa o desvio-padrão das amostras de predição (LI *et al.*, 2018; HU *et al.*, 2019).

Por fim, empregou-se o teste da região elíptica de confiança conjunta (EJCR) no conjunto de predição para examinar a existência de viés (bias) e, portanto, avaliar a precisão dos modelos (ORMOND; FERNANDES; DINIZ, 2022).

3.4.3. Softwares

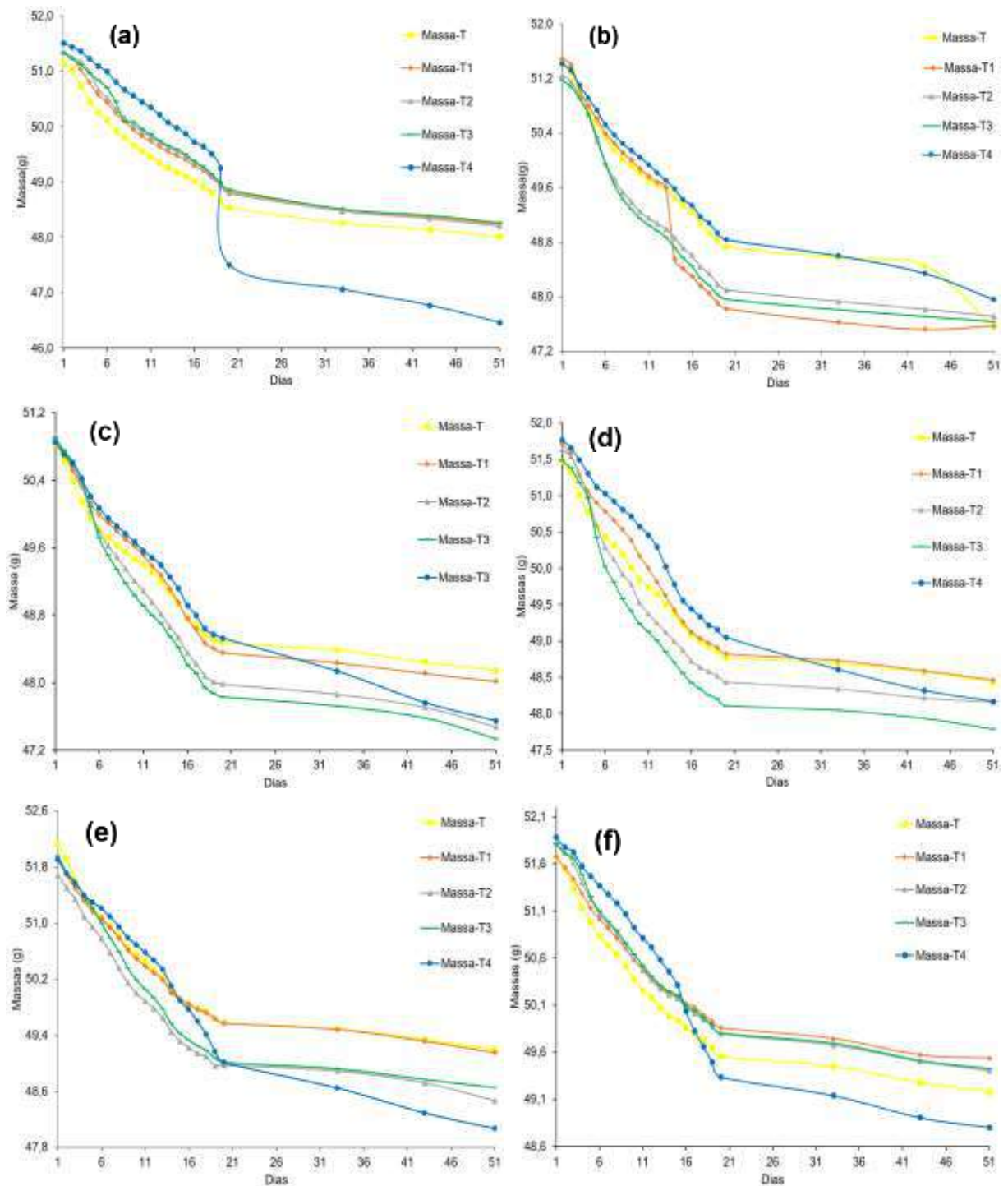
Para a construção dos modelos PCA foi realizada utilizando o PLS toolbox®, disponível em <https://eigenvector.com/software/pls-toolbox/> e os modelos ACO-LDA, PLS, iPLS e iSPA-PLS foram desenvolvidos em rotinas feitas em laboratório, ambos em ambiente Matlab®.

4. RESULTADOS E DISCUSSÃO

4.1. Caracterização físico-química dos extratos processados de *Agave* spp.

O estudo da perda de massa dos genótipos de *Agave* spp. está ilustrado na **Figura 11**. O conservante utilizado tem como função evitar os diferentes tipos de deterioração, principalmente impedir a fermentação dos extratos (MELO, GONÇALVES, SOARES, 2005). Os tratamentos contendo a maior concentração do conservante (azul escuro) mantiveram uma menor taxa de perda de massa (VELÁSQUEZ *et al.*, 2009) até próximo aos 20 dias de análise. Observamos um decréscimo brusco no valor das massas de duas amostras, em azul escuro (**Figura 11a**) e em laranja (**Figura 11b**), tais variações estão relacionadas a derramamento de parte das amostras durante o período dos experimentos.

Figura 11 - Taxa de degradação para os seis genótipos de *Agave* spp. **(a)** A, **(b)** B, **(c)** C, **(d)** D, **(e)** E e **(f)** F no período de 51 dias, em função da dose do conservante.



Massa – T = 0 (m/ v) % de conservante, Massa – T1 = 0,01 (m/ v)% de conservante, Massa – T2 = 0,05 (m/ v)% de conservante, Massa – T3 = 0,1 (m/ v)% de conservante, Massa – T4 = 0,5 (m/ v)% de conservante.

Fonte: Elaborado pelo autor, 2023.

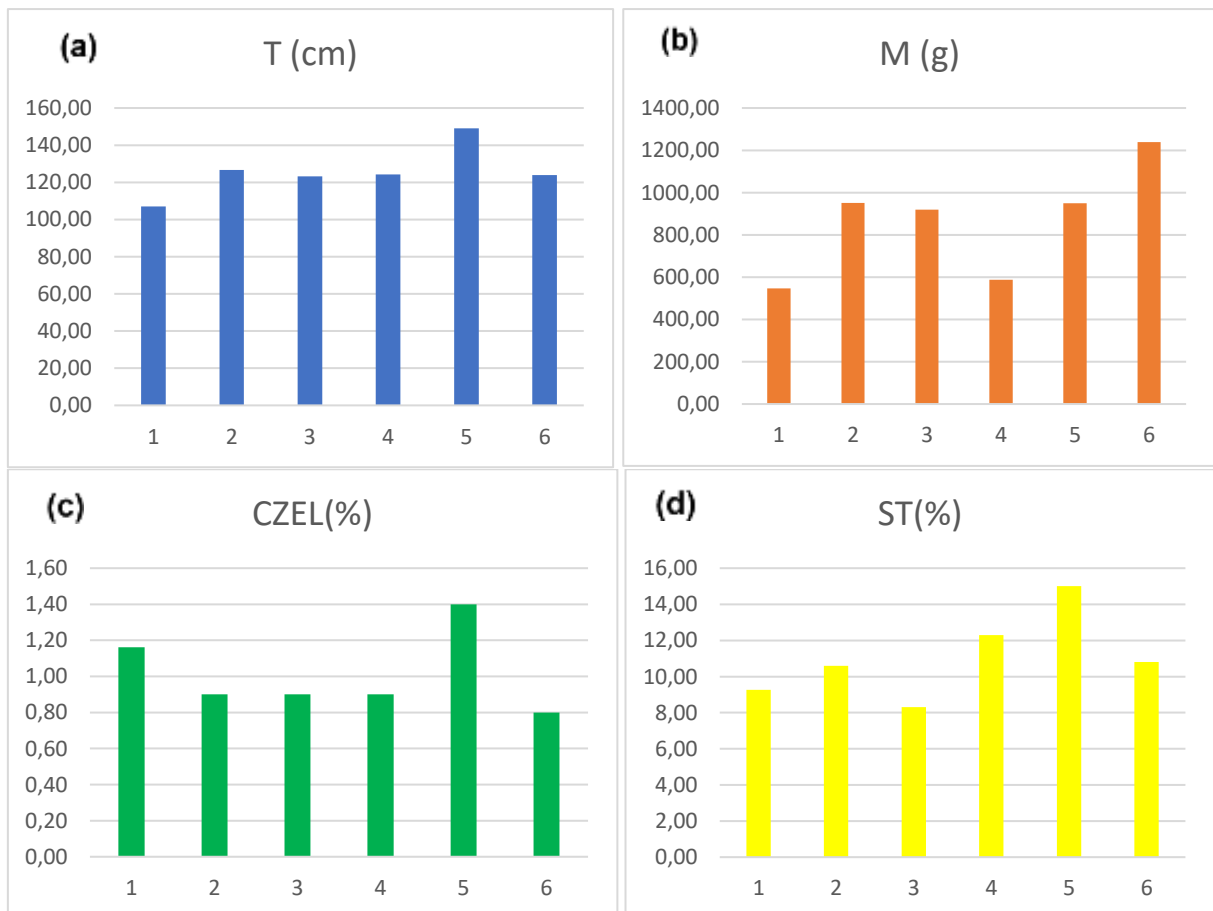
Os resultados obtidos para as caracterizações físico-químicas dos seis genótipos de *Agave* spp. estudados estão dispostos na **Tabela 2** em anexo.

Uma variação de 546,6 g a 1238,6 g na massa das folhas dos genótipos foi evidenciada. Sendo os genótipos F e B os que forneceram maior rendimento de massa verde. Em relação ao comprimento das folhas, verificou-se uma variação de 107,1 cm a 149,1 cm. Os genótipos E e B forneceram os maiores valores de comprimento. Os valores de sólidos totais (ST) variou entre 8,3% a 15%. Sendo os menores valores médios observados para o genótipo C e A.

Os genótipos que apresentam valores mais expressivos de massa (M) e comprimento (T) das folhas e baixos teores de sólidos totais (ST) e teor de água (TA) são mais promissores para a produção de fibras. De forma que os genótipos F e B seriam os melhores para a produção de fibra.

Em relação a composição mineral representada pelo teor de cinzas (CZEL), foi observada uma variação de 0,8% a 1,4%. Sendo os maiores valores em média apresentados para os genótipos E e A.

Figura 12 - Variação (a) comprimento (T em cm), (b) massa (M em g) das folhas (c) teor de cinzas (CZEL%) e (d) sólidos totais (ST%) no extrato e



Fonte: Elaborado pelo autor, 2023.

Já os genótipos que tendem a maior rendimento para a produção da fração líquida são aqueles que apresentam maior TA e ST, como é o caso dos genótipos C e E.

A qualidade da fração líquida obtida dos genótipos de *Agave* spp. foi medida pela acidez decorrente da fermentação espontânea do material, que ocorre em poucos minutos após a coleta. Os genótipos D e F forneceram os menores valores de acidez natural (2,8 a 3,0 mg NaOH/ mL), quando comparado aos demais genótipos, forneceram valores de acidez entre 4,2 a 5,1 mg NaOH/ mL.

Uma avaliação estatística utilizando o teste de Tukey (**Figura 13**) foi realizada a fim de verificar semelhanças entre pares de genótipos. De acordo com a avaliação dos valores de confiança, pode-se afirmar que quando o valor 0 (zero) não está contido nos intervalos de confiança, pode-se concluir que as médias (genótipos) são significativamente diferentes ao nível de probabilidade de 95% confiança (**Tabela 2**).

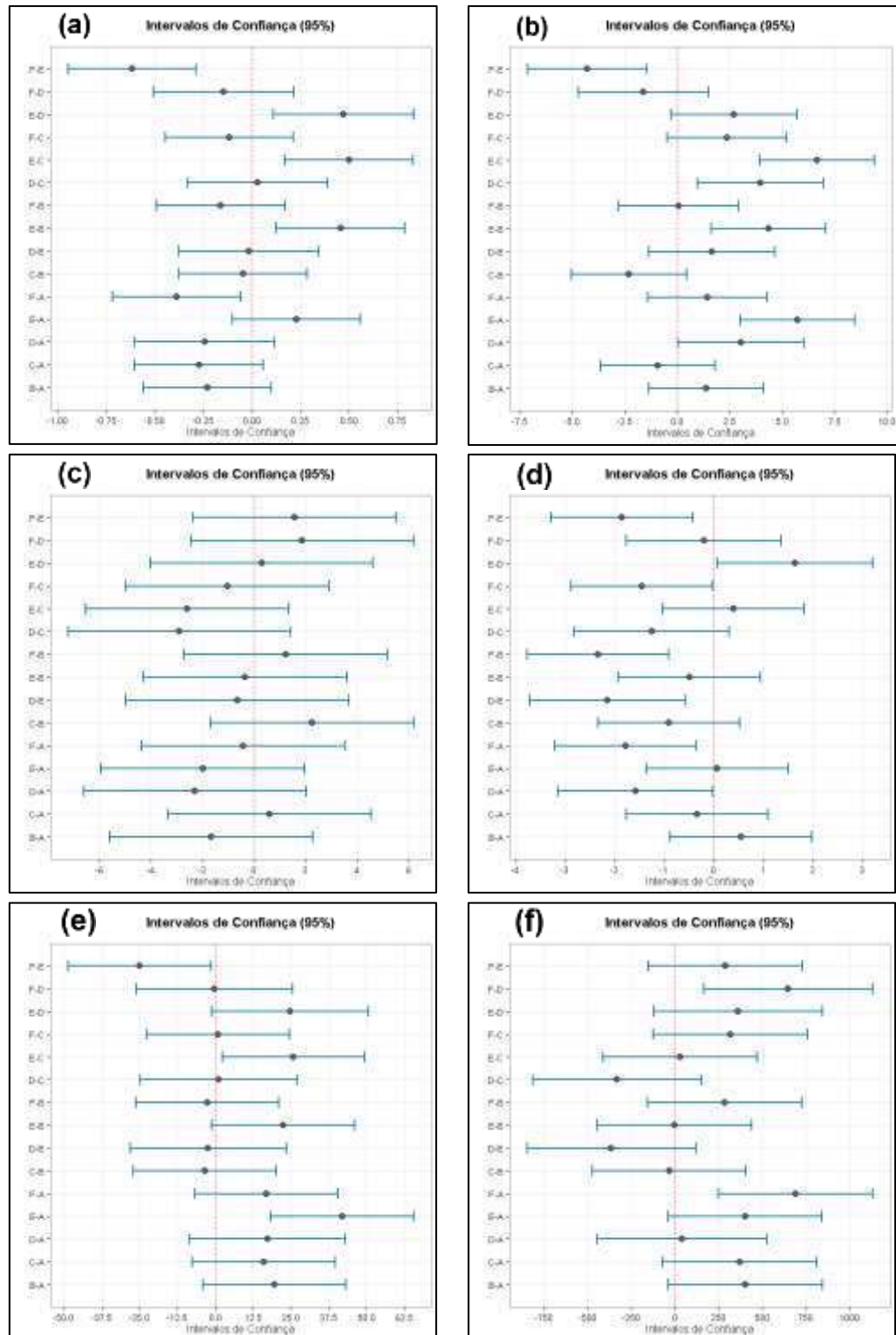
De acordo como a **Tabela 2** e **Figura 13** os genótipos são estatisticamente semelhantes em termos do teor de água. No entanto são estatisticamente diferentes em termos de teor de cinzas, acidez total e sólidos totais. Em termos de massa os genótipos A, D e F são estatisticamente diferentes (B, C e E são semelhantes) e em termos de comprimento A, C e E são estatisticamente diferentes (B, D e F são semelhantes).

Tabela 2 - Genótipos estatisticamente diferentes de acordo com o teste Tukey.

CZEL(%)	ST(%)	TA(%)	AT (mg NaOH.mL ⁻¹)	T(cm)	M(g)
F(0,8) - E(1,4)	F(10,8) - E(15,0)	-	F(2,8) - E(4,6)	E(149,1) - C(123,2)	F(1238,6) - D(587,6)
E(1,4) - D(0,9)	E(15,0) - C(8,3)	-	E(4,6) - D(3,0)	E(149,1) - A(107,1)	F(1238,6) - A(546,6)
E(1,4) - C(0,9)	D(12,3) - C(8,3)	-	F(2,8) - C(4,2)	-	-
E(1,4) - B(0,9)	E(15,0) - B(10,6)	-	F(2,8) - B(5,1)	-	-
F(0,8) - A(1,2)	E(15,0) - A(9,3)	-	D(3,0) - B(5,1)	-	-
-	D(12,3) - A(9,3)	-	F(2,8) - A(4,5)	-	-
-	-	-	D(3,0) - A(4,5)	-	-

Fonte: Elaborado pelo autor, 2023.

Figura 13 - Teste de Tukey. **(a)** CZEL: Cinzas Extrato Líquido, **(b)** ST: Sólidos Totais, **(c)** TA: Teor Água, **(d)** AT: Acidez Titulável, **(e)** T: Tamanho médio e **(f)** M: Massa foliar média.



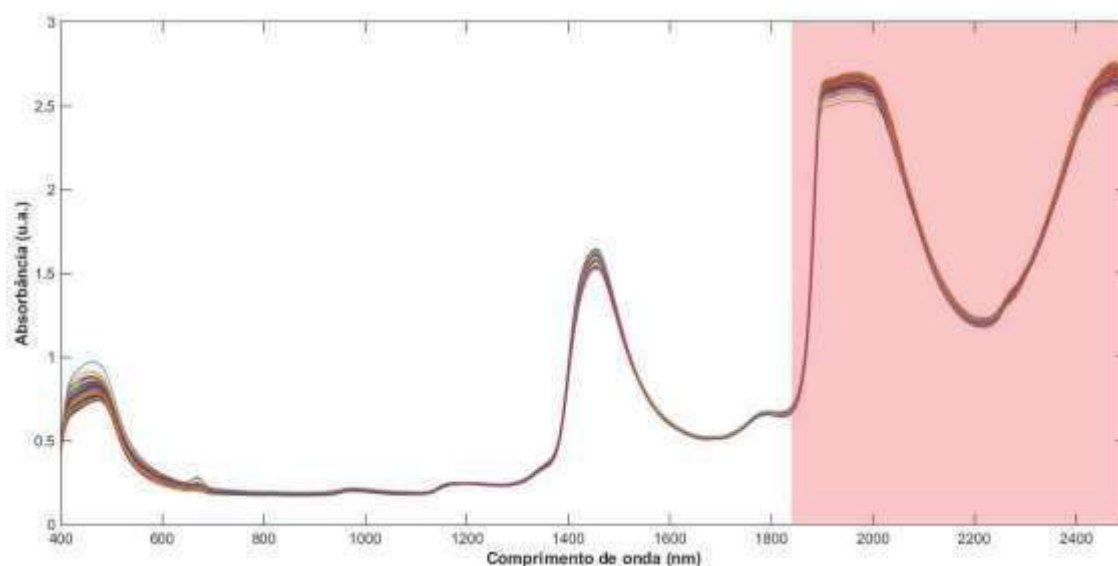
Fonte: Elaborado pelo autor, 2023.

4.2. Modelos baseados na espectroscopia VIS-NIR

4.2.1. Reconhecimento de Padrões

Os espectros de transfectância de 400 a 2500 nm convertidos em unidades contínuas, podem ser visualizados na **Figura 14** para o conjunto de todas as 200 amostras da fração líquida dos genótipos de *Agave* spp. É observável que na faixa compreendida entre 1837 nm e 2500 nm em vermelho, nota-se um perfil bastante ruidoso dos espectros e um valor de absorbância superior a 2 u.a, que não seria adequado para a construção de modelos quimiométricos. Por esse motivo, essa região do espectro foi removida.

Figura 14 - Espectros brutos dos genótipos de *Agave* spp. completos com faixa removida destacada em vermelho.

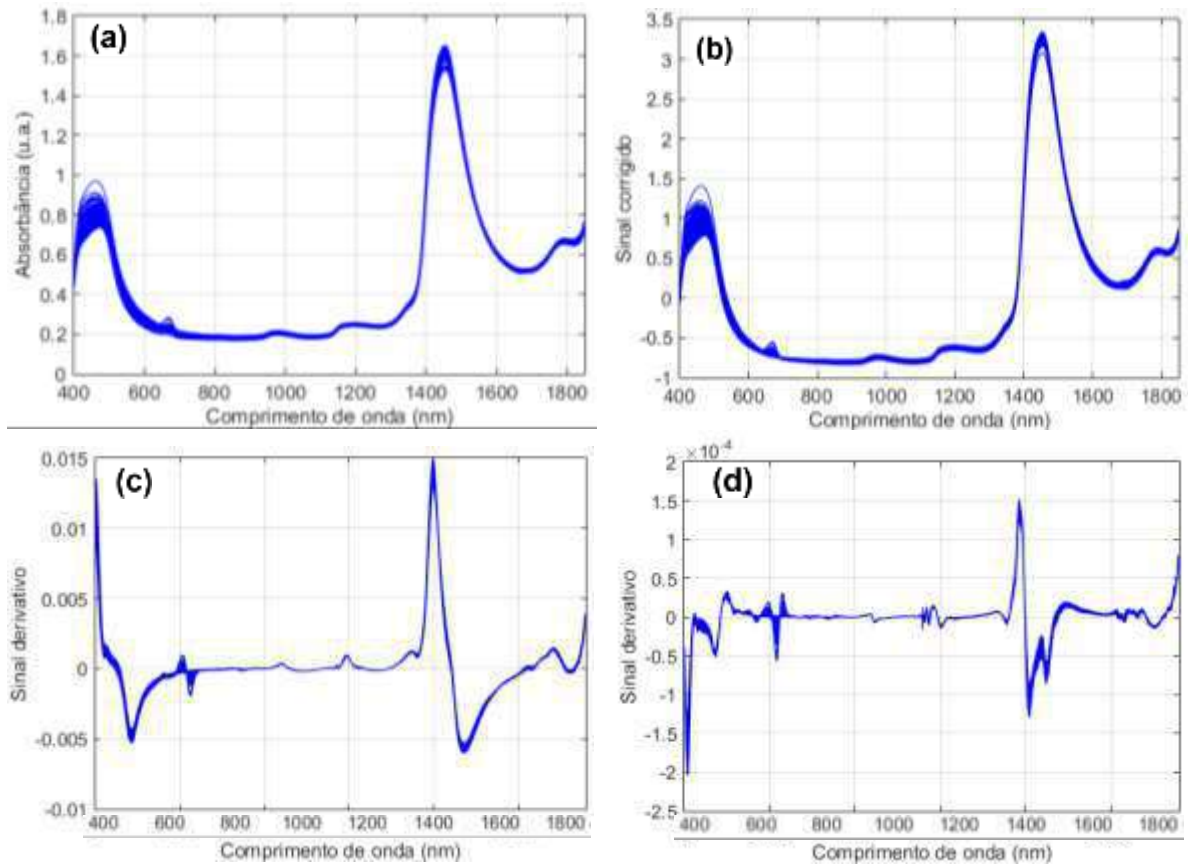


Fonte: Elaborado pelo autor, 2023.

Após a seleção da faixa de trabalho, foram aplicados pré-processamento para a construção dos modelos de reconhecimento de padrões. Na **Figura 15** são visualizados a faixa de trabalho sem pré-processamento e com a aplicação do pré-processamento correção normal padrão (SNV), primeira e segunda derivada Savitzky Golay com polinômio de 1ª e 2ª ordem com janelas de 5, 7 e 15.

Como os resultados, ou seja visualização dos agrupamentos, foi similar para os pré-processamentos optou-se por adotar o SNV (**Figura 16**) para a construção das PCA's.

Figura 15 - (a) faixa de trabalho, (b) dados pré-processados com SNV e (c) 1ª derivada com polinômio de 1ª ordem e (d) 2ª derivada com polinômio de 2ª ordem.

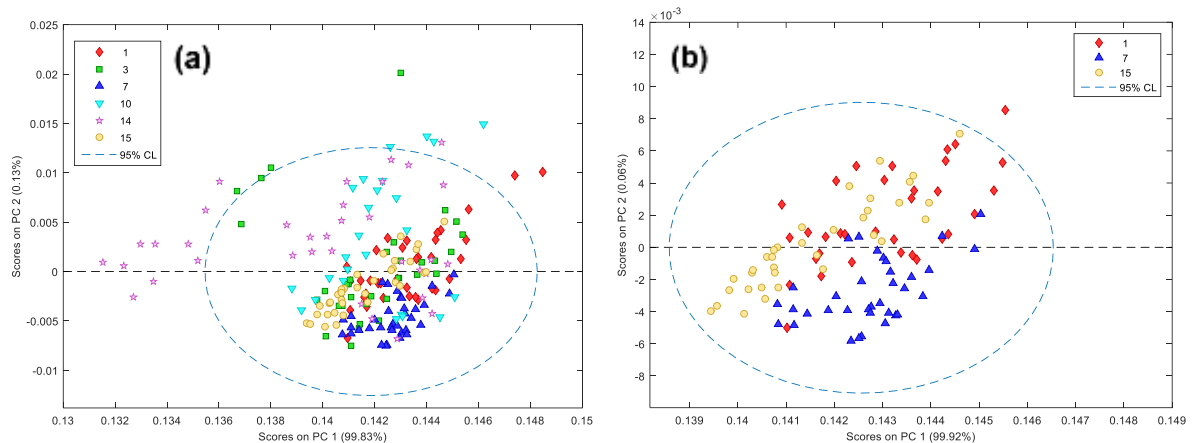


Fonte: Elaborado pelo autor, 2023.

A PCA construída com os espectros dos 6 genótipos não permitiu observar tendência de separação entre os genótipos de acordo com os gráficos de escores (**Figura 16a**). No entanto, observou-se uma tendência de separação entre os genótipos 1, 7 e 15 (A, C e F de acordo com a sessão 3.2). Os resultados para os demais genótipos evidenciaram sobreposição e dispersão, considerando o gráfico dos escores.

Deste modo, construiu-se uma nova PCA apenas com os genótipos 1, 7 e 15. Em que foi observada uma melhor tendência de separação (**Figura 16b**) entre os três genótipos. O modelo foi construído com 3 PC's com 99,99% de variância explicada.

Figura 16 - Gráfico de escores para amostras de *Agave* spp. na região VIS-NIR. **(a)** todas os genótipos e **(b)** para os genótipos 1, 7 e 15 (A, C e F).



Fonte: Elaborado pelo autor, 2023.

Deste modo, procedeu-se o desenvolvimento do modelo de reconhecimento de padrões supervisionado com seleção de variáveis (ACO-LDA). Inicialmente, tentou-se construir o modelo discriminatório para cinco classes de genótipos como observa-se no gráfico da **Figura 17a**. A classe D (genótipo 10) foi deixada de fora, devido a grande dispersão entre as amostras da classe. No entanto, evidenciou-se baixa especificidade e sensibilidade para algumas classes como demonstrado na **Tabela 3**. O poder discriminante e as variáveis selecionadas pelo algoritmo ACO são ilustradas na **Figura 17b**. Neste caso o conjunto de treinamento foi composto por 120 amostras e o de teste por 60 amostras.

Em seguida foi avaliada a capacidade preditiva do modelo ACO-LDA construído apenas para os genótipos 1,7 e 15 (**Tabela 4**). Neste caso, foi alcançada uma boa distinção entre as classes, como pode ser observado na **Figura 17 c e d**. A boa separação entre as classes pode ser observada na **Figura 17c** e as variáveis selecionadas pelo ACO-LDA com maior poder discriminante se encontram na região de 1500 à 1600 nm representados na **Figura 17 d)** para o modelo com três classes. Para a construção do modelo ACO-LDA o conjunto de treinamento foi composto por 60 amostras e o de teste por 45 amostras.

Tabela 3 - Figuras de mérito para o modelo ACO-LDA para as 5 classes de genótipos de *Agave* spp.

ACO-LDA toda faixa e todas as classes SNV						
	Etapa de treinamento			Etapa de teste		
% Classificação correta	82,49%			83,08%		
% Classificação incorreta	17,50%			16,91%		
Erros Totais	37			22		
	Precisão	Sensibilidade	Especificidade	Precisão	Sensibilidade	Especificidade
Classe 1	0,73	0,90	0,93	0,93	0,93	0,98
Classe 2	0,58	0,61	0,90	0,50	0,60	0,94
Classe 3	0,61	0,92	0,94	0,74	0,74	0,89
Classe 4	0,97	0,63	0,99	0,79	0,92	0,96
Classe 5	0,43	0,56	0,88	0,58	0,58	0,87

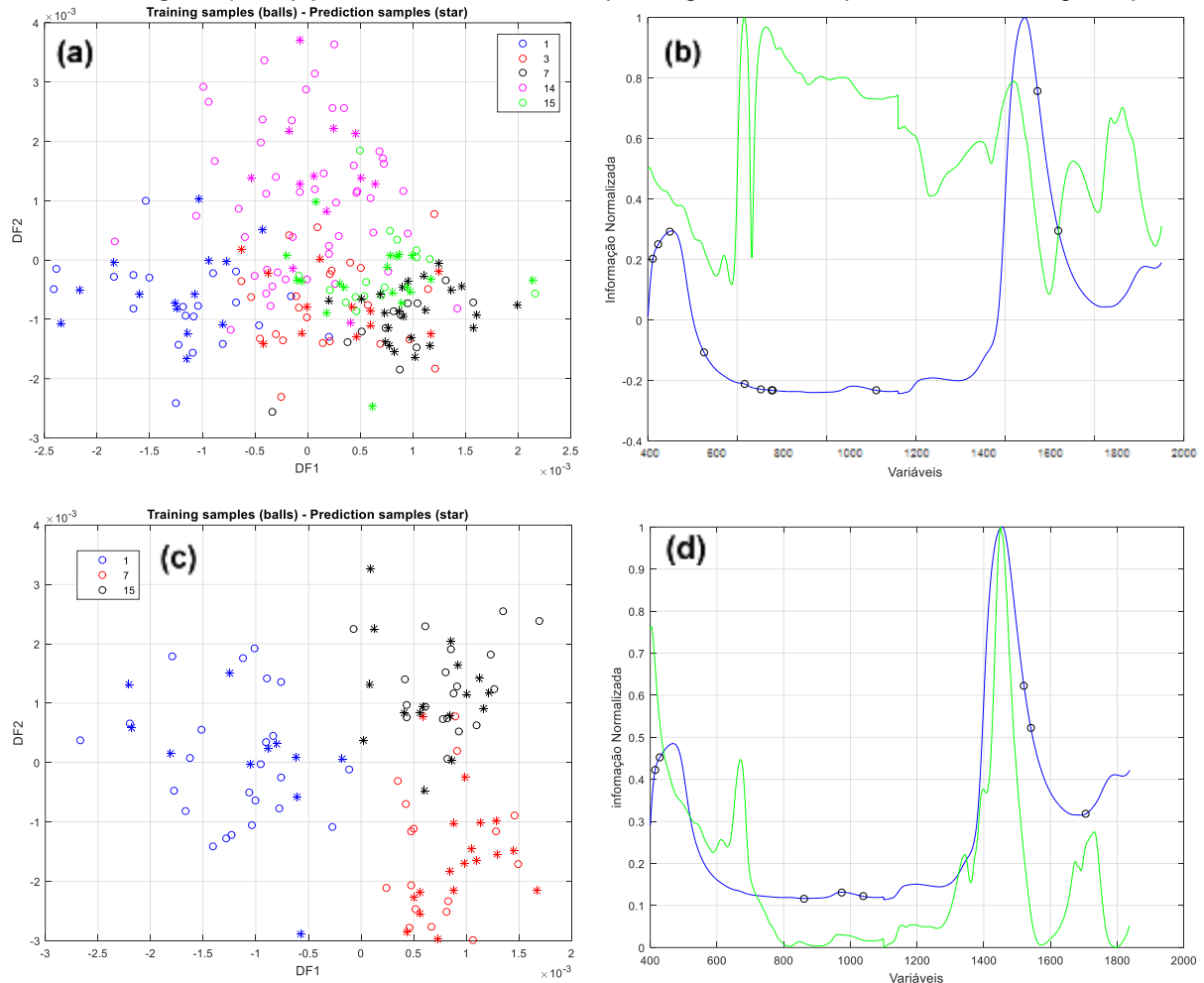
Fonte: Elaborado pelo autor, 2023.

Tabela 4 - Figuras de mérito para o modelo ACO-LDA para as 3 classes de genótipos de *Agave* spp.

ACO-LDA toda faixa e todas as classes SNV						
	Etapa de treinamento			Etapa de teste		
% Classificação correta	96,14%			94,70%		
% Classificação incorreta	3,85%			5,29%		
Erros Totais	3			3		
	Precisão	Sensibilidade	Especificidade	Precisão	Sensibilidade	Especificidade
Classe 1	1,00	0,96	1,00	1,00	0,91	1,00
Classe 2	0,94	0,88	0,98	0,89	0,94	0,93
Classe 3	0,90	1,00	0,95	0,94	0,94	0,97

Fonte: Elaborado pelo autor, 2023.

Figura 17 - (a) Modelo ACO-LDA com SNV para as 5 classes de genótipos de *Agave* spp. **(b)** Variáveis selecionadas pelo algoritmo ACO para as 5 classes de genótipos. **(c)** Modelo ACO-LDA com SNV para 3 classes de genótipos. **(d)** Variáveis selecionadas pelo algoritmo ACO para 3 classes de genótipos.



Fonte: Elaborado pelo autor, 2023.

4.2.2. Modelos de Regressão (PLS) e Seleção de Variáveis (iSPA PLS e iPLS)

Os modelos de calibração multivariada utilizando os espectros VIS-NIR para a predição da acidez no extrato líquido foi realizada com os intervalos selecionados pelo iPLS e iSPA-PLS (GOMES, 2012) e os resultados foram comparados com o modelo construído com o espectro completo. Foram construídos modelos com os dados tratados com SNV, 1ª derivada com polinômio Savitzky Golay e 2ª derivada com polinômio Savitzky Golay de 2ª ordem com janelas de 5, 7 e 15 pontos. As figuras de mérito para os modelos construídos estão identificadas na **Tabela 5**. O conjunto de dados foi dividido em conjunto de calibração (130 espectros) e predição (70 espectros).

O melhor modelo PLS foi construído com 8 variáveis latentes, que explicam 82,42% dos dados espectrais (X) e 89,40% da variação da acidez (mg NaOH/ mL) (Y), construído com os dados tratados com 2ª derivada com janela de 5 pontos e polinômio de 2ª ordem. Apesar dos modelos construídos com os demais pré-processamentos testados terem apresentado menores valores de RMSE e REP, os mesmos necessitaram de um alto número de variáveis latentes, o que pode levar a um sobreajuste dos modelos construídos, além de baixa robustez.

De acordo com as figuras de mérito evidenciadas, os modelos que apresentaram menor REP foram o iSPA-PLS com os dados tratados com 2ª derivada 5 Pontos e polinômio de 2ª ordem e o modelo iSPA-PLS com os dados tratados com 1ª derivada 5 Pontos e polinômio de 1ª ordem. A maioria dos modelos construídos forneceram coeficiente de correlação próximo a unidade, demonstrando uma relação ideal entre os espectros e a acidez total. Os valores de RMSE também foram adequados, sendo menores que cerca de 20% do limite inferior da curva de calibração. No entanto, pode-se observar que alguns modelos forneceram discrepância entre os valores de RMSECV e RMSEP, sugerindo, em alguns casos a possibilidade de sobreajuste. A quantidade de variáveis latentes e variáveis espectrais necessárias para a construção dos modelos foi considerada alta, no entanto considera-se aceitável devido à complexidade do problema estudado.

Tabela 5 - Resultados dos modelos PLS, iSPA-PLS e iPLS obtidos para determinação da acidez dos genótipos de *Agave* spp.

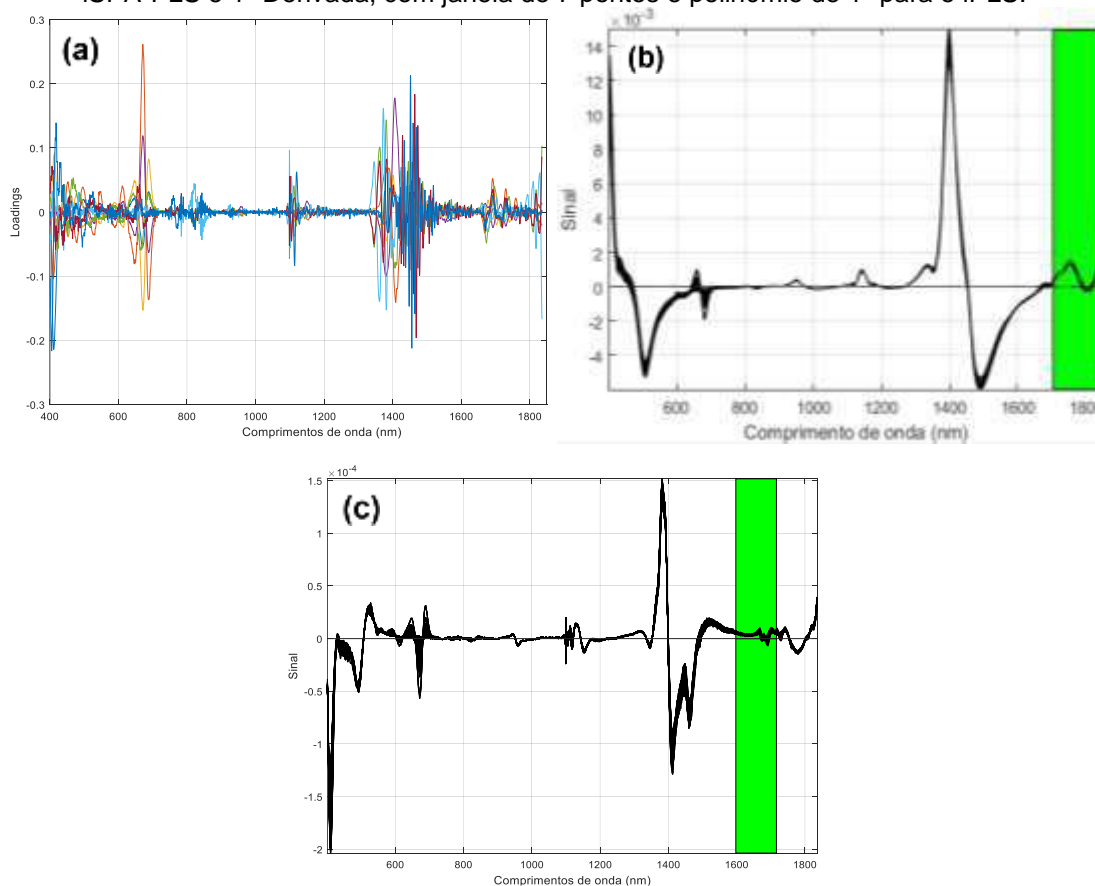
Modelo	Pré-processamentos	Outliers	Figuras de mérito					
			Fatores PLS (intervalos selecionados)	RMSECV	r	RMSEP	r _{Pred}	REP
PLS	1ª Derivada 5 Pontos	0	16	0,58	0,978	0,981	0,830	9,483
	1ª Derivada 7 Pontos	0	17	0,54	0,981	0,753	0,911	7,277
	1ª Derivada 15 Pontos	0	29	0,20	0,997	0,723	0,915	6,958
	2ª Derivada 5 Pontos	0	8	1,62	0,705	1,594	0,545	15,45
	SNV	1	19	1,09	0,918	0,712	0,916	6,901
iPLS	1ª Derivada 5 Pontos	0	7	1,00	0,927	0,987	0,834	9,545
	1ª Derivada 7 Pontos	1	7 (1)	1,00	0,925	0,870	0,873	8,408
	1ª Derivada 15 Pontos	2	8	1,04	0,922	0,867	0,860	8,358
	SNV	1	9	1,09	0,916	0,966	0,865	9,402
iSPA PLS	1ª Derivada 5 Pontos	0	14	0,57	0,988	0,911	0,860	8,801
	1ª Derivada 7 Pontos	0	17	0,37	0,991	0,766	0,907	7,405
	1ª Derivada 15 Pontos	0	29	0,20	0,998	0,716	0,917	6,901
	2ª Derivada 5 Pontos	0	8 (1)	1,43	0,846	1,110	0,771	10,765
	SNV	0	13	0,62	0,976	0,616	0,932	7,000

Fonte: Elaborado pelo autor, 2023.

Diante do exposto, tendo em vista as figuras de mérito evidenciadas e o princípio da parcimônia, foram escolhidos como melhores modelos o iPLS e iSPA-PLS construídos com 1ª e 2ª derivada, janela de 7 e 5 pontos e polinômios de 1ª e 2ª ordem, respectivamente.

Na **Figura 18 a)** é apresentado o gráfico do *loadings* para o modelo construído com PLS, as regiões próximas a 700 nm, 1400 nm e 1500 nm são as regiões mais informativas para o modelo, já na **Figura 18 b)** iPLS e **c)** iSPA-PLS estão representadas as variáveis selecionadas para ambos modelos, a região entre 1600 à 1750 nm e de 1750 à 1850 nm, respectivamente, é mais informativa. Na literatura é possível encontrar autores que citam bandas encontradas na região selecionada que podem ser atribuídas a presença de compostos fenólicos como flavonoides e taninos condensados, são elas 1415 – 1512 nm, 1650 – 1750 nm (TOLEDO-MARTÍN *et al.*, 2018; DYKES *et al.*, 2014; MENEZES, MARDER, COSTA, 2013).

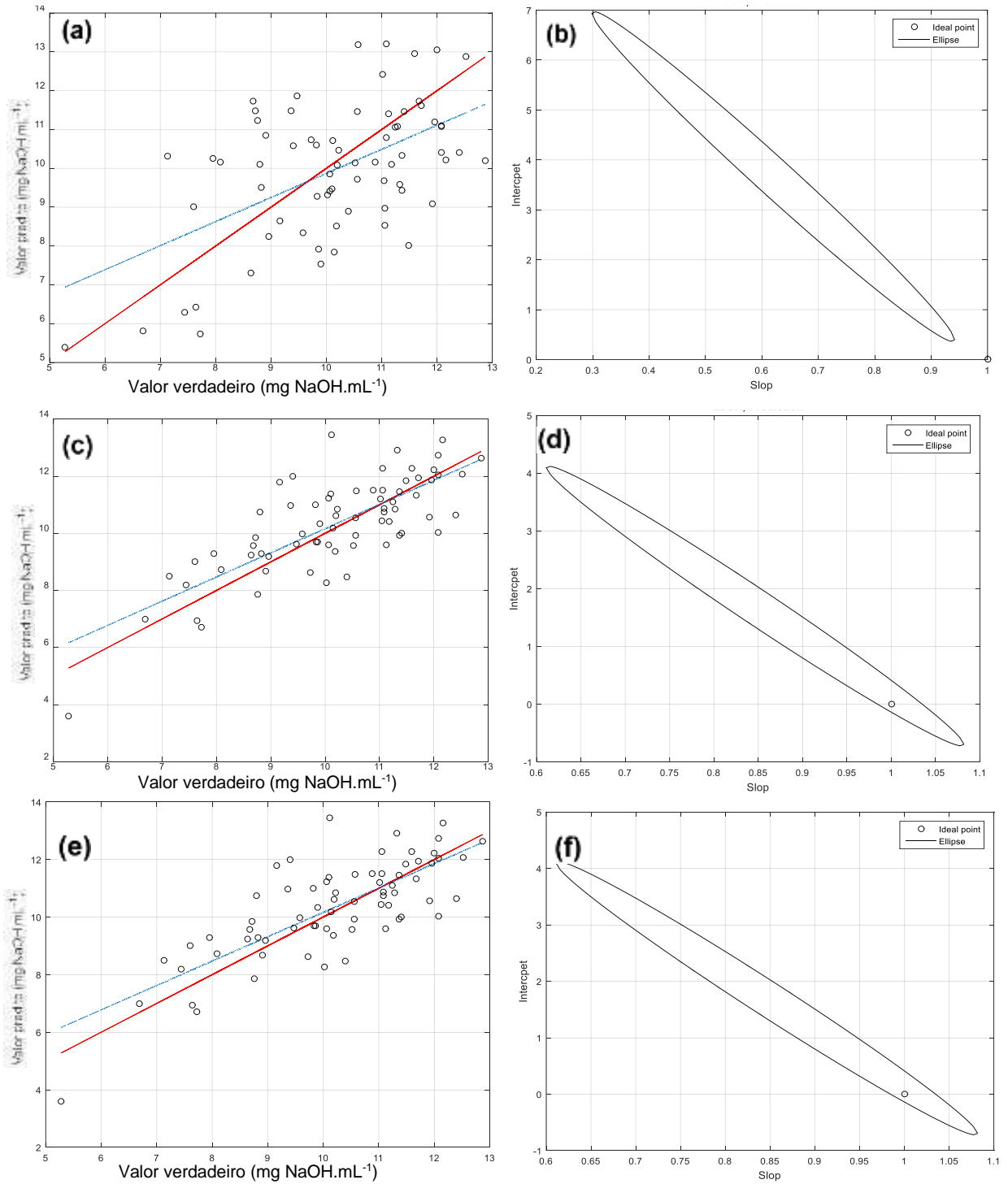
Figura 18 - Gráficos de Loadings do modelo **(a)** PLS as variáveis selecionadas para os modelos **(b)** iPLS e **(c)** iSPA-PLS, utilizando 2ª Derivada, com janela de 5 pontos e polinômio de 2º para o PLS e iSPA-PLS e 1ª Derivada, com janela de 7 pontos e polinômio de 1º para o iPLS.



Fonte: Elaborado pelo autor, 2023.

Na **Figura 19 a)** PLS, **c)** iPLS e **e)** iSPA-PLS está ilustrado os gráficos dos valores preditos versus referência para a determinação acidez dos genótipos de *Agave* spp. Além disso, na **Figura 19 b)** PLS, **d)** iPLS e **f)** iSPA-PLS são apresentados os gráficos da região conjunta da elipse de confiança. Pode-se observar que para o modelo PLS o ponto teórico ideal não se encontra dentro da elipse, indicando a baixa exatidão do modelo construído. Entretanto, para os modelos iPLS e iSPA-PLS não há presença de erro sistemática, devido a distribuição das amostras por ambos os lados da linha de bissecção e pode-se observar a proximidade entre a bissetriz (em vermelho) e o segmento de reta para a regressão (em azul) obtida pelo modelo e a distribuição das amostras tendendo a um comportamento linear, confirmado nos gráficos EJCRC na **Figura 19 d)** e **f)**.

Figura 19 - Gráficos de valor predito versus referência obtido para a determinação da acidez pelos modelos **(a)** PLS e **(c)** iPLS **(e)** iSPA-PLS. Gráficos de EJCR com as elipses de confiança contendo seus respectivos pontos teóricos ideais obtidos para a determinação da acidez pelos modelos **(b)** PLS **(d)** iPLS e **(f)** iSPA-PLS, utilizando 2ª Derivada, com janela de 5 pontos e polinômio de 2º para o PLS e iSPA-PLS e 1ª Derivada, com janela de 7 pontos e polinômio de 1º para o iPLS.



Fonte: Elaborado pelo autor, 2023.

Os modelos propostos apresentaram um bom desempenho para classificação dos genótipos de *Agave* spp. estudados. Observou-se que a capacidade preditiva dos modelos é influenciada pelos pré-processamentos espectrais utilizados. O número de variáveis latentes utilizados para construção dos modelos está relacionado a complexidade do problema avaliado.

5. CONCLUSÕES

Os genótipos estudados foram avaliados em termos de características físico químicas. Os extratos com maior rendimento foram observados para os genótipos C e E. De acordo com o teste de Tukey não há diferença significativa entre os genótipos em relação ao teor de água, em razão de estarem nas mesmas condições edafoclimáticas. A menor acidez total para os extratos foi observada para os genótipos D e F.

Diante do exposto, podemos afirmar, em prova de conceito, o uso da espectroscopia vis-NIR aliada a análise discriminante linear e ao algoritmo bioinspirado ACO é capaz de distinguir entre alguns genótipos de *Agave* spp. Além disso, os modelos de calibração multivariada foram adequados para a determinação rápida da acidez total no extrato líquido do *Agave* spp.

Em relação a análise exploratória, utilizando PCA, os espectros VIS-NIR, pode-se afirmar que os genótipos A, C e F (1, 7 e 15) são diferentes em sua composição com características intrínsecas para maior aprofundamento. As demais classes de genótipos são bastante dispersas e sobrepostas.

Em relação aos modelos de reconhecimento de padrões supervisionados utilizando LDA e o algoritmo ACO para seleção de variáveis, foi alcançada uma taxa de classificação superior a 90%. Evidenciando uma boa capacidade do modelo em distinguir três classes de genótipos.

Os modelos de calibração multivariada construídos com os espectros VIS-NIR e os algoritmos iPLS e iSPA-PLS forneceram um bom desempenho. Sendo os modelos iPLS e iSPA-PLS construídos com os dados pré-processados 1ª e 2ª derivada, janela de 7 e 5 pontos e polinômios de 1ª e 2ª ordem, respectivamente.

6. PERSPECTIVAS

- Utilizar o algoritmo bioinspirado firefly para prever a acidez das amostras de extrato de sisal.
- Construir modelos multivariados baseados em dados NIR para predição de propriedades como açúcares naturais e metabólitos como taninos, triterpenos e saponinas em extratos de sisal.
- Predição da toxicidade do extrato de sisal utilizando NIR e quimiometria.

REFERÊNCIAS

ABEBAYEHU, Seyoum Getaneh; ENGIDA, Adam Mekonnen. Preparation of Biocomposite Material with Superhydrophobic Surface by Reinforcing Waste Polypropylene with Sisal (*Agave sisalana*) Fibers. **International Journal of Polymer Science**. v. 2021, p. 1-15, 2021.

AKARACHANTACHOTE, Noppamas *et al.* Cutoff threshold of variable importance in projection for variable selection. **Int J Pure Appl Math**, v. 94, n. 3, p. 307-322, 2014.

ALVES, Maria Odete; SANTIAGO, Eduardo Girão. Tecnologia e relações sociais de produção no setor sisaleiro nordestino. **Revista Econômica do Nordeste**, v. 37, n. 3, p. 368-381, 2006.

ANDRADE, Robson; ORNELAS, Jackson; BRANDÃO, Weliton. Situação atual do sisal na Bahia e suas novas possibilidades de utilização e aproveitamento. **Comunicação SEAGRI**, p. 14-19, 2011.

ARAÚJO, Mário César Ugulino *et al.* The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. **Chemometrics and intelligent laboratory systems**, v. 57, n. 2, p. 65-73, 2001.

BARRETO, Stella Maria Andrade Gomes. **Utilização do subproduto do beneficiamento do sisal (*Agave sisalana* Perrine): desenvolvimento de nanoemulsões cosméticas e avaliação da segurança e eficácia**. 2017. 117f. Dissertação (Mestrado em Ciências Farmacêuticas) – Universidade Federal do Rio Grande do Norte, Natal, 2017.

BAUTISTA-MONTES, Erika; HERNÁNDEZ-SORIANO, Laura; SIMPSON, June. Advances in the micropropagation and genetic transformation of *Agave* species. **Plants**, v. 11, n. 13, p. 1757, 2022.

BEĆ, Krzysztof B.; HUCK, Christian W. Breakthrough potential in near-infrared spectroscopy: Spectra simulation. A review of recent developments. **Frontiers in chemistry**, v. 7, n. 48, p. 1-22, 2019.

BOVENS, Michael *et al.* Chemometrics in forensic chemistry - Part I: implications to the forensic workflow. **Forensic science international**, v. 301, p. 82-90, 2019.

BRAGA, V. A. *et al.* **Composição inseticida de extrato de *Agave* híbrida para o combate ao mosquito *Aedes Aegypti* em qualquer uma de suas fases de vida**. 2018, Brasil. Patente: Privilégio de Inovação. Número do registro: BR1020180130056, Instituição de registro: INPI – Instituto Nacional da Propriedade Industrial. Depósito: 25/06/2018.

BRERETON, Richard G. *et al.* Chemometrics in analytical chemistry - part I: history, experimental design and data analysis tools. **Analytical and Bioanalytical Chemistry**, v. 409, p. 5891-5899, 2017.

BROWN, Karen. Agave sisalana perrine. **University of Florida Center for Aquatic and Invasive Plants**, v. 7922, p. 18-21, 2002.

CHAPMAN, James. *et al.* Combining chemometrics and sensors: Toward new applications in monitoring and environmental analysis. **Chemical Reviews**, v. 120, n. 13, p. 6048-6069, 2020.

CHEN, Hui *et al.* Rapid determination of cotton content in textiles by near-infrared spectroscopy and interval partial least squares. **Analytical Letters**, v. 51, n. 17, p. 2697-2709, 2018.

CUI, Hai-Feng *et al.* Automatic and rapid discrimination of cotton genotypes by near infrared spectroscopy and chemometrics. **Journal of Analytical Methods in Chemistry**, v. 2012, p. 1-7, 2012.

DAMASCENO, Josilda. Cavalcante Amorim *et al.* Resíduo de sisal sobre *Meloidogyne javanica* em plantas de alface. **Enciclopédia biosfera**, v.14 n. 26, p. 641-650, 2017.

DIAS, Leonardo de Souza *et al.* Incorporação de resíduos da produção de fibras de sisal em argamassa: Efeitos nas propriedades físicas e mecânicas. **Revista Matéria**, v. 26, n. 3, 2021.

DINIZ, Paulo Henrique Gonçalves Dias. Chemometrics-assisted color histogram-based analytical systems. **Journal of Chemometrics**, v. 34, n. 12, p. 1-27, 2020.

DUPONT, Madeleine F. *et al.* Chemometrics for environmental monitoring: a review. **Analytical Methods**, v. 12, n. 38, p. 4597-4620, 2020.

DYKES, Linda *et al.* Prediction of total phenols, condensed tannins, and 3-deoxyanthocyanidins in sorghum grain using near-infrared (NIR) spectroscopy. **Journal of cereal science**, v. 60, n. 1, p. 138-142, 2014.

EMPRESA BRASILEIRA DE PESQUISA E AGROPECUÁRIA. **Sisal**. Disponível em: <https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos/sisal>. Acesso em: 21 de fev. 2023.

ESBENSEN, Kim H. *et al.* **Multivariate data analysis: in practice: an introduction to multivariate data analysis and experimental design**. 5ª ed. Multivariate Data Analysis, 2002.

FOOD AND AGRICULTURE ORGANIZATION. Future Fibres: Sisal. Roma, 2023. Disponível em: <https://www.fao.org/economic/futurefibres/fibres/sisal/en/>. Acesso em 03 Ago. 2023

FERNANDES, Janaina Rodrigues *et al.* Nanopartículas de sílica silanizada como compatibilizante em compósitos de fibras de sisal/polietileno. **Polímeros**. v. 27, n. especial, p. 61-69, 2017.

FERREIRA, Daniel da Silva. **Marcadores espectrais no visível e infravermelho próximo para identificação de sementes em progênies de *Ricinus communis* L.** 2015. 135 f. Tese (Doutor em Agronomia). Universidade Federal da Paraíba, Areia, 2015.

FERREIRA, Márcia Miguel Castro. **Quimiometria: conceitos, métodos e aplicações.** Editora da UNICAMP, 2015.

FERREIRA, Márcia Miguel Castro. Quimiometria III-Revisitando a análise exploratória dos dados multivariados. **Química Nova**, v. 45, p. 1251-1264, 2022.

FISHER, Ronald A. The use of multiple measurements in taxonomic problems. **Annals of eugenics**, v. 7, n. 2, p. 179-188, 1936.

GAUB, M. *et al.* Analysis of plant extracts by NIRS: simultaneous determination of kavapyrones and water in dry extracts of Piper methysticum Forst. **Journal of pharmaceutical and biomedical analysis**, v. 36, n. 4, p. 859-864, 2004.

GARCÍA-MENDOZA, Abisaí Josué; FRANCO MARTÍNEZ, Irma Sonia; SANDOVAL GUTIÉRREZ, Daniel. Cuatro especies nuevas de Agave (Asparagaceae, Agavoideae) del sur de México. **Acta Botanica Mexicana**, n. 126, p. 1-18, jan. 2019.

GEBREMICAEL, Tesfay Gebretsadkan; MOHAMED, Y. A.; VAN DER ZAAG, Pieter. Attributing the hydrological impact of different land use types and their long-term dynamics through combining parsimonious hydrological modelling, alteration analysis and PLSR analysis. **Science of the Total Environment**, v. 660, p. 1155-1167, 2019.

GOMES, Adriano de Araújo. **Algoritmo das Projeções Sucessivas aplicado à seleção de variáveis em regressão PLS.** 2012. 124 f. Dissertação (Mestrado em Química). Universidade Federal da Paraíba, Departamento de Química, João Pessoa, 2012.

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

GONDIM, Tarcísio Marcos de Souza *et al.* Prospecção de genótipos de agave para obtenção de suco para bioinseticida. In: SILVA-MATOS, Raissa Rachel Salustriano da; MORAES, Lídia Ferreira; SILVA, Fabíola Luzia de Souza. Desenvolvimento da pesquisa científica, tecnologia e inovação na agronomia 3. **Atena**. p. 129-137, 2022.

GSCHAEDLER, Anne Christine. *et al.* **Panorama de aprovechamiento de los agaves em México.** 1ª ed. Guadalajara: Agared, 2017.

HONORATO, Fernanda Araújo. **Previsão de propriedades das gasolinas do nordeste empregando espectroscopia NIR/MIR e transferência de calibração.** 2006. 90 f. Tese (Doutorado em Química). Universidade Federal de Pernambuco, Departamento de Química Fundamental, Recife, 2006.

HU, Rong *et al.* Optimization of soluble solids content prediction models in 'Hami'melons by means of Vis-NIR spectroscopy and chemometric tools. **Infrared Physics & Technology**, v. 102, p. 102999, 2019.

INSTITUTO ADOLFO LUTZ, I. A. L. **Métodos físico-químicos para análise de alimentos**. 4ª ed. Instituto Adolf Lutz, p.1020, 2008.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Levantamento sistemático da produção agrícola: Pesquisa Mensal de Previsão e Acompanhamento das Safras Agrícolas no Ano Civil**. 2020. Disponível em: <https://sidra.ibge.gov.br>. Acesso em: 03 ago. 2023.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Produção Agrícola: Lavoura permanente (Sisal)**. 2021. Disponível em: <https://cidades.ibge.gov.br/brasil/pesquisa/15/12040>. Acesso em: 09 ago. 2023.

JAMRÓGIEWICZ, Marzena. Application of the near-infrared spectroscopy in the pharmaceutical technology. **Journal of pharmaceutical and biomedical analysis**, v. 66, p. 1-10, 2012.

KHERIF, Ferath; LATYPOVA, Adeliya. Principal component analysis. In: **Machine Learning**, p. 209-225, 2020.

KUMAR, Raj; SHARMA, Vishal. Chemometrics in forensic science. **TrAC Trends in Analytical Chemistry**, v. 105, p. 191-201, 2018.

LEE, Loong Chuen; JEMAIN, Abdul Aziz. On overview of PCA application strategy in processing high dimensionality forensic data. **Microchemical Journal**, v. 169, p. 1-16, 2021.

LI, Wenbing *et al.* Quantitative analysis of calorific value of coal based on spectral preprocessing by laser-induced breakdown spectroscopy (LIBS). **Energy & Fuels**, v. 32, n. 1, p. 24-32, 2018.

LÓPEZ-HIDALGO, Cristina; MENÉNDEZ, Macarena; JORRIN-NOVO, Jesús V. Phytochemical composition and variability in *Quercus ilex* acorn morphotypes as determined by NIRS and MS-based approaches. **Food Chemistry**, v. 338, p. 127803, 2021.

MANLEY, Marena. Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. **Chemical Society Reviews**, v. 43, n. 24, p. 8200-8214, 2014.

MARTIN, Adriana R. *et al.* Caracterização química e estrutural de fibra de sisal da variedade *Agave sisalana*. **Polímeros**, v. 19, n. 1, p. 40-46, 2009.

MARTINS, José Antônio; SARGENTELLI, Vagner. Infravermelho próximo na avaliação quantitativa foliar. **Prospectus**. v. 3, n. 1, p. 33-55, 2021.

MEDINA, Júlio César. O sisal. **Secretaria da Agricultura do Estado de São Paulo**, São Paulo, 1954.

MELLADO-MOJICA, Erika; LÓPEZ, Mercedes Guadalupe. Identification, classification, and discrimination of agave syrups from natural sweeteners by infrared spectroscopy and HPAEC-PAD. **Food Chemistry**, v. 167, p. 349-357, 2015.

MELO, Nathália Ramos de; GONÇALVES, Maria Paula Junqueira C.; SOARES, Nilda de Fátima Ferreira. Nisina: um conservante natural para alimentos. **Revista Ceres**, v. 52, n. 304, 2005.

MENEZES, Caren Machado; MARDER, Luciano; COSTA, Adilson Ben da. Utilização da NIRS para a determinação de tanino e de umidade diretamente na casca de *Acacia mearnsii* para a indústria de extração. In: **Workshop em Sistemas e Processos Industriais**. 2013.

MILANEZ, Karla Danielle Tavares Melo; PONTES, Márcio José Coelho. Classification of edible vegetable oil using digital image and pattern recognition techniques. **Microchemical Journal**, v. 113, p. 10-16, 2014.

MILANEZ, Karla Danielle Tavares de Melo. **Seleção de variáveis robustas para transferência de modelos de classificação empregando o algoritmo das projeções sucessivas**. 2017. 131 f. Dissertação (Mestrado em Química). Universidade Federal da Paraíba, Departamento de Química, João Pessoa, 2017.

MISHRA, Sidharth Prasad. *et al.* Multivariate statistical data analysis-principal component analysis (PCA). **International Journal of Livestock Research**, v. 7, n. 5, p. 60-78, 2017.

MOREIRA, Alessandro César de Oliveira. **Espectroscopia NIR, CG-EM e quimiometria para o controle de qualidade do óleo de copaíba (*Copaifera spp.*)**. 2018. 171 f. Tese (Doutorado em Tecnologias Química e Biológica). Universidade de Brasília, Instituto de Química, Brasília, 2018.

MOREIRA, J.; BELTRÃO, N. E de M.; SILVA, O. R. R. F. da. Botânica e morfologia do sisal. In: SILVA, Odilon Reny Ribeiro Ferreira da; BELTRÃO, Napoleão Esberard de Macêdo. **Agronegócio do sisal no Brasil**. Serviço de Produção de Informação, p. 25-34, 1999.

NUNES, Fabíola da Cruz *et al.* Estudo da atividade larvicida da *Agave sisalana* contra *Aedes Aegypti*. 2013. In: SILVA, Odilon Reny Ribeiro Ferreira da; BELTRÃO, Napoleão Esberard de Macêdo. **Agronegócio do sisal no Brasil**. Serviço de Produção de Informação, p. 25-34, 1999.

OLIVEIRA, Louise Helena Guimarães de *et al.* *Agave sisalana* extract induces cell death in *Aedes aegypti* hemocytes increasing nitric oxide production. **Asian Pacific Journal of Tropical Biomedicine**, v. 6, n. 5, p. 396-399, 2016.

OLIVIERI, Alejandro C. **Introduction to multivariate calibration: A practical approach**. Springer, 2018.

ORMOND, Tayrine de Souza; FERNANDES, David Douglas de Sousa; DINIZ, Paulo Henrique Gonçalves Dias. An eco-friendly analytical methodology based on digital images for quality control of commercial *Mikania glomerata* syrups. **Microchemical Journal**, v. 178, p. 107338, 2022.

OSBORNE, Brian G. Near-infrared spectroscopy in food analysis. **Encyclopedia of analytical chemistry: applications, theory and instrumentation**, p. 1-14, 2006.

OZAKI, Yukihiro. *et al.* **Near-infrared spectroscopy: theory, spectral analysis, instrumentation, and applications**. Springer, 2021.

PASQUINI, Celio. Near infrared spectroscopy: A mature analytical technique with new perspectives—A review. **Analytica chimica acta**, v. 1026, p. 8-36, 2018.

PASQUINI, Celio. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. **Journal of the Brazilian chemical society**, v. 14, n. 2, p. 198-219, 2003.

PONTES, Aline S. *et al.* Ant colony optimization for variable selection in discriminant linear analysis. **Journal of Chemometrics**, v. 34, n. 12, p. e3292, 2020.

QUEIROGA, Vicente de Paula *et al.* SISAL (*Agave sisalana*, Perrine) TECNOLOGIAS DE PLANTIO E UTILIZAÇÃO. **Associação da Revista Eletrônica A Barriguda**, p. 1-217, 2021.

RAYA, Fábio Trigo. **Expressão gênica de " Agave" para bioenergia: " Agave" gene expression for bioenergy**. 2018. 91 f. Dissertação (Mestre em Genética e Biologia Molecular). Universidade Estadual de Campinas, Campinas, 2018.

RIJAL, Deepa *et al.* Quality estimation of *Agave tequilana* leaf for bioethanol production. **Journal of Near Infrared Spectroscopy**, v. 24, n. 5, p. 453-465, 2016.

ROBERT, Manuel L. *et al.* Wild and agronomically important *Agave* species (Asparagaceae) show proportional increases in chromosome number, genome size, and genetic markers with increasing ploidy. **Botanical Journal of the Linnean Society**, v. 158, n. 2, p. 215-222, 2008.

SAEYS, Wouter *et al.* Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review. **Postharvest Biology and Technology**, v. 158, p. 1-19, 2019.

SAFO, Sandra E.; AHN, Jeongyoun. General sparse multi-class linear discriminant analysis. **Computational Statistics & Data Analysis**, v. 99, p. 81-90, 2016.

SANTANA, Felipe Bachion 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, Daniele Oliveira Justo dos; FONTES, Cintia Maria Ariani; LIMA, Paulo Roberto Lopes. Uso de agregado miúdo reciclado em matrizes cimentícias para compósitos reforçados com fibras de sisal. **Revista Matéria**, v. 22, nº 1, 2017.

SANTOS, Edinusia Moreira Carneiro; SILVA, Onildo Araujo da. SISAL IN BAHIA - BRAZIL. **Mercator**, v. 16, p. 1-13, 2017.

SANTOS, Jener David Gonçalves *et al.* Chemicals from Agave sisalana biomass: isolation and identification. **International Journal of Molecular Sciences**, v. 16, n. 4, p. 8761-8771, 2015.

SANTOS, Lucinéia dos *et al.* **Método de obtenção de extrato fermentado de Agave Sisalana Perrine (sisal), solução aquosa obtida e uso da solução para controle de ácaros de culturas agrícolas.** p. 1-13, 2022.

SANTOS, Maria Betania Hermenegildo dos. **Fenotipagem não destrutiva usando espectroscopia no infravermelho próximo e quimiometria em sementes de mamona.** 2013. 77 f. Tese (Doutorado em Química). Universidade Federal da Paraíba, João Pessoa, 2013.

SAUZIER, Georgina. *et al.* Chemometrics in forensic science: approaches and applications. **Analyst**, v. 146, n. 8, p. 2415-2448, 2021.

SAVASTANO JUNIOR, Holmer; PIMENTEL, Lia Lorena. Viabilidade do aproveitamento de resíduos de fibras vegetais para fins de obtenção de material de construção. **Revista brasileira de engenharia agrícola e ambiental**, v. 4, p. 103-110, 2000.

SIESLER, Heinz W. *et al.* **Near-infrared spectroscopy: principles, instruments, applications.** John Wiley & Sons, 2008.

SILVA, Francisca Célia; RIBEIRO, Alan Bezerra; RIBEIRO, P. R. S. Avaliação da qualidade de plantas medicinais comercializadas no município de Imperatriz, Maranhão. **Scientia Plena**, v. 13, n. 2, p. 1-9, 2017.

SILVA, J. T. *et al.* Rapadura de caldo de cana pura e com adição de açúcar refinado: análise físico-química e classificação por espectroscopia NIRR e quimiometria. **Revista Virtual de Química**, v. 11, n. 4, p. 1318-1329, 2019.

SILVA, Odilon Reny Ribeiro Ferreira da *et al.* **Cultivo do Sisal no Nordeste Brasileiro.** EMBRAPA, 2008.

SILVA, Rafael Mota da *et al.* Controle de *Aspergillus welwitschiae* e a podridão vermelha com resíduo líquido do desfibramento das folhas de sisal. **Ciência Agrícola**, v. 17, n. 3, p. 13-21, 2019.

SILVA, Wilson Castro *et al.* Avaliação do efeito tóxico de extratos de *Palicourea marcgravii* St. Hil. (Rubiaceae) sobre *Aetalion* sp. (Hemiptera: Aetalionidae) em laboratório. **Revista Brasileira de Biociências**, v. 7, n. 2, p. 129-133, 2009.

SIMÕES, Simone da Silva. **Desenvolvimento de métodos validados para a determinação de captopril usando espectrometria NIR e calibração multivariada**. 2008. 83 f. Tese (Doutor em Química). Universidade Federal da Paraíba, João Pessoa, 2008.

SISAL Brasil. **Rural Thomaz**, 2023. Disponível em: <http://thomazrural.com.br/wp-content/uploads/2023/01/Sisal-Brasil-Informativo-2022.pdf>. Acesso em: 15 de jun. 2023.

SKOOG, Douglas. A. *et al.* **Fundamentos de Química Analítica**. Cengage Learning, 2009.

SKOOG, Douglas. A.; HOLLER, F. James.; NIEMAN, Timothy. A. **Fundamentos de química analítica**. Reverté, 2020.

SOARES, Liz F. *et al.* Avaliação de espectrômetro NIR portátil e PLS-DA para a discriminação de seis espécies similares de madeiras amazônicas. **Química Nova**, v. 40, p. 418-426, 2017.

SOARES, José Hildemarcio Mendes; ARRUDA, Danilo Raimundo de; AMARANTE, Patrícia Araújo. Transformações tecnológicas e econômicas do sisal no Nordeste do Brasil. **Research, Society and Development**, v. 11, n. 5, 2022.

SOARES, Sófacles Figueredo Carreiro *et al.* The successive projections algorithm. **TrAC Trends in Analytical Chemistry**, v. 42, p. 84-98, 2013.

SOUZA, André Marcelo de *et al.* Experimento didático de quimiometria para calibração multivariada na determinação de paracetamol em comprimidos comerciais utilizando espectroscopia no infravermelho próximo: um tutorial, parte II. **Química Nova**, v. 36, p. 1057-1065, 2013.

SOUZA, Géssica Oliveira de. **Avaliação da Atividade Antimicrobiana e Perfil Químico do Extrato de Sisal (*Agave Sisalana Perrine*)**. 2017. 48 f. Trabalho de Conclusão de Curso (Graduação em Farmácia). Faculdade Maria Milza, Governador Mangabeira, 2017.

SUINAGA, Fábio Akiyoshi., SILVA, Odilon Reny Ribeiro Ferreira da., COUTINHO, Wirtton Macedo. **Cultivo de Sisal na Região Semi-Árida do Nordeste Brasileiro**. Campina Grande, PB, 2006.

TEIXEIRA, Kelly Sivocy Sampaio. **Uso da quimiometria na determinação simultânea do teor dos fármacos em comprimido com dose fixa combinada empregado no tratamento de tuberculose**. 2017. 83 f. Tese (Doutorado em Desenvolvimento e Inovação Tecnológica em Medicamentos). Universidade Federal do Rio Grande do Norte, Faculdade de Farmácia, Rio Grande do Norte, 2017.

THOMPSON, James M. **Infrared spectroscopy**. Jenny Stanford Publishing, 2018.

TIAN, Han *et al.* Weighted SPXY method for calibration set selection for composition analysis based on near-infrared spectroscopy. **Infrared Physics & Technology**, v. 95, p. 88-92, 2018.

TOLEDO-MARTÍN, Eva María *et al.* Quantification of total phenolic and carotenoid content in blackberries (*Rubus Fruticosus* L.) using near infrared spectroscopy (NIRS) and multivariate analysis. **Molecules**, v. 23, n. 12, p. 3191, 2018.

TREJO-TORRES, Jorge C.; GANN, George D.; CHRISTENHUSZ, Maarten J. M. The Yucatan Peninsula is the place of origin of sisal (*Agave sisalana*, Asparagaceae): historical accounts, phytogeography and current populations. **Botanical Sciences**, vol.96, n.2, p. 366-379, jun. 2018.

VELÁSQUEZ, Paula Andrea Toro *et al.* Cinética da fermentação e taxas de degradação de forrageiras tropicais em diferentes idades de corte estimadas pela técnica de produção de gases in vitro. **Revista Brasileira de Zootecnia**, v. 38, p. 1695-1705, 2009.

VIDAL, Mariane C.; SALDANHA, Rodolfo; VERISSIMO, Mario Alvaro Aloisio. Bioinsumos: o programa nacional e a sua relação com a produção sustentável. In: GINDRI, Diego Medeiros, MOREIRA, Patrícia Almeida Barroso, VERISSIMO, Mario Alvaro Aloisio. Sanidade vegetal: uma estratégia global para eliminar a fome, reduzir a pobreza, proteger o meio ambiente e estimular o desenvolvimento econômico sustentável. **CIDASC**, p. 382-409, 2020.

VILVERT, João Claudio *et al.* Avaliação não destrutiva de atributos de qualidade em maçãs e peras usando espectroscopia no infravermelho próximo e quimiometria. **Revista Brasileira de Fruticultura**, v. 45, p. e-969, 2023.

WULANDARI, Lesty; PERMANA, Bayu Dwi; KRISTININGRUM, Nia. Determination of total flavonoid content in medicinal plant leaves powder using infrared spectroscopy and chemometrics. **Indonesian Journal of Chemistry**, v. 20, n. 5, p. 1044-1051, 2020.

XIAOBO, Zou. *et al.* Variables selection methods in near-infrared spectroscopy. **Analytica chimica acta**, v. 667, n. 1-2, p. 14-32, 2010.

XU, Yi *et al.* Raman spectroscopy coupled with chemometrics for food authentication: A review. **TrAC Trends in Analytical Chemistry**, v. 131, p. 1-13, 2020.

YAO, Siyu *et al.* A novel handheld FT-NIR spectroscopic approach for real-time screening of major cannabinoids content in hemp. **Talanta**, v. 247, p. 123559, 2022.

ZHU, Fa *et al.* Neighborhood linear discriminant analysis. **Pattern Recognition**, v. 123, p. 108422, 2022.

ZULLO, Marco Antônio Teixeira *et al.* Efeito de diferentes condições de fermentação sobre o teor e composição da fração de sapogeninas do suco de sisal. **Bragantia**, v. 43, p. 479-486, 1984.

APÊNDICES

Tabela A1 - Dados físico-químicos de genótipos de *Agave* spp. da Embrapa Algodão.

.....77

Tabela A1 - Dados físico-químicos de genótipos de *Agave* spp. da Embrapa Algodão.

Amostra Genótipo	CZEL(%)*	ST(%)*	TA (%)*	AT (mg NaOH/ mL)*	T (cm)*	M (g)*
A_PI1	1,13	10,47	84,57	5,53	105,7	775,4
A_PI2	1,27	8,89	85,54	5,19	112,7	511,3
A_PI3	1,08	10,08	82,82	2,20	100	376,04
A_PI4	1,25	8,39	83,01	2,49	102,3	415,7
A_PI5	1,01	8,48	86,56	4,49	108,3	602,6
A_PI6	1,45	9,72	83,75	7,08	109,7	484,5
A_PI7	0,94	8,77	84,62	4,75	110,7	660,5
Média A	1,2	9,3	84,4	4,5	107,1	546,6
B_PI1	0,78	10,39	83,98	4,69	133,3	1276,7
B_PI2	1,02	10,75	82,89	5,64	120	782,8
B_PI3	1,01	9,59	85,59	5,27	129	930,3
B_PI4	0,89	9,51	83,99	5,36	108,7	715,32
B_PI5	0,99	11,4	74,12	5,08	115,3	423
B_PI6	0,76	11,36	84,52	4,87	146,7	1254,5
B_PI7	1,08	11,42	84,22	4,77	133	1280,1
Média B	0,9	10,6	82,8	5,1	126,6	951,8
C_PI1	0,84	9,92	83,16	4,81	117,3	762,7
C_PI2	0,77	8,96	85,13	4,73	132,7	1182,4
C_PI3	0,82	7,59	84,91	3,59	117,7	874,7
C_PI4	0,92	7,39	86,42	4,20	126	954
C_PI5	0,89	7,22	86,75	3,31	132,7	1071,7
C_PI6	1,01	8,38	85,01	4,24	116,7	813
C_PI7	0,97	8,86	83,71	4,56	119	775,9
Média C	0,9	8,3	85,0	4,2	123,2	919,2
D_PI3	0,87	13,12	79,39	2,61	142	549,4
D_PI4	1,02	12,56	81,99	3,30	140	698,1
D_PI5	1,1	13,36	79,24	3,63	133,3	553
D_PI6	0,59	11,9	82,73	2,41	133	855,7
D_PI7	1,01	10,54	87,26	2,86	73	281,9
Média D	0,9	12,3	82,1	3,0	124,3	587,6
E_PI1	1,15	9,5	85,76	4,29	149	970,7
E_PI2	1,13	17,33	78,98	4,86	141,7	1165,2
E_PI3	1,39	16,54	83,08	4,50	145,7	758,6
E_PI4	1,53	13,12	86,13	4,76	114,7	475,1
E_PI5	1,77	14,75	80,50	5,37	167	955,4
E_PI6	1,59	13,35	81,94	4,55	160,3	818,2
E_PI7	1,17	20,34	80,57	3,96	165	1502,5
Média E	1,4	15,0	82,4	4,6	149,1	949,4
F_PI1	0,34	11,47	84,72	1,69	137,3	1860,9
F_PI2	0,77	10,28	82,52	2,85	125	1115,9
F_PI3	1,3	9,87	83,36	4,01	106,3	699,2
F_PI4	0,95	10,29	84,26	3,76	124	983,38
F_PI5	0,78	11,7	84,42	2,48	128	1327,6
F_PI6	0,52	10,53	84,83	2,13	128,7	1266,5
F_PI7	0,75	11,25	83,84	2,46	118,7	1416,5
Média F	0,8	10,8	84,0	2,8	124,0	1238,6

Legenda: CZEL: Cinzas Extrato Líquido; ST: Sólidos Totais; TA: Teor Água; AT: Acidez Titulável; T: Tamanho médio; M: Massa foliar média.

Fonte: Elaborado pelo autor, 2023.