



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE CIÊNCIAS AGRÁRIAS**  
**DEPARTAMENTO DE FITOTECNIA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA/FITOTECNIA**

**ERIVANESSA COSTA SOUSA SARMENTO**

**ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO E ANÁLISE**  
**MULTIVARIADA NA AVALIAÇÃO DO POTENCIAL FISIOLÓGICO DE**  
**SEMENTES DE GERGELIM E SORGO**

**FORTALEZA**

**2022**

ERIVANESSA COSTA SOUSA SARMENTO

ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO E ANÁLISE MULTIVARIADA  
NA AVALIAÇÃO DO POTENCIAL FISIOLÓGICO DE SEMENTES DE GERGELIM E  
SORGO

Tese apresentada ao Programa de Pós-Graduação em Agronomia/Fitotecnia da Universidade Federal do Ceará, como parte dos requisitos para a obtenção do grau de Doutora em Agronomia/Fitotecnia. Área de concentração: Manejo de Agroecossistemas.

Orientador: Prof. Dr. Alek Sandro Dutra

Coorientadora: Prof<sup>a</sup>. Dra. Livia Paulia Dias Ribeiro

FORTALEZA

2022

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

- S255e Sarmiento, Erivanessa Costa Sousa.  
Espectroscopia no infravermelho próximo e análise multivariada na avaliação do potencial fisiológico de sementes de gergelim e sorgo / Erivanessa Costa Sousa Sarmiento. – 2022.  
63 f. : il. color.
- Tese (doutorado) – Universidade Federal do Ceará, Centro de Ciências Agrárias, Programa de Pós-Graduação em Agronomia (Fitotecnia), Fortaleza, 2022.  
Orientação: Prof. Dr. Alek Sandro Dutra.  
Coorientação: Profa. Dra. Lívia Paulia Dias Ribeiro.
1. Sesamum indicum. 2. Sorghum bicolor. 3. PLS-DA. 4. NIR. 5. Quimiometria. I. Título.  
CDD 630
-

ERIVANESSA COSTA SOUSA SARMENTO

ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO E ANÁLISE MULTIVARIADA  
NA AVALIAÇÃO DO POTENCIAL FISIOLÓGICO DE SEMENTES DE GERGELIM E  
SORGO

Tese apresentada ao Programa de Pós-Graduação em Agronomia/Fitotecnia da Universidade Federal do Ceará, como parte dos requisitos para a obtenção do grau de Doutora em Agronomia/Fitotecnia. Área de concentração: Manejo de Agroecossistemas.

Aprovada em 27 de setembro de 2022.

BANCA EXAMINADORA

---

Prof. Dr. Alek Sandro Dutra (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof<sup>a</sup>. Dra. Lívia Paulia Dias Ribeiro (Coorientadora)  
Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB)

---

Prof<sup>a</sup>. Dra. Charline Zaratín Alves  
Universidade Federal de Mato Grosso do Sul (UFMS)

---

Prof. Dr. Márcio Dias Pereira  
Universidade Federal do Rio Grande do Norte (UFRN)

---

Prof. Dr. Allan Nilson de Sousa Dantas  
Instituto Federal do Rio Grande do Norte (IFRN)

A minha avó, Maria José, por todo esforço e dedicação à minha criação.

Aos meus pais, Benedito e Arivanilza, e ao meu marido, Fernando Sarmiento, por todo apoio e incentivo durante minha vida acadêmica.

## AGRADECIMENTOS

A Deus, pela vida, pelo amparo durante os momentos difíceis, e pela sabedoria para tomar as decisões mais importantes da minha vida;

À Universidade Federal do Ceará (UFC), por meio do Programa de Pós-graduação em Agronomia/Fitotecnia (PPGAF), pela oportunidade de realização do curso de Doutorado;

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001;

Ao professor e orientador, Dr. Alek Sandro Dutra, pela orientação e confiança;

A professora e coorientadora, Dra. Livia Paulia Dias, pela orientação da realização deste trabalho, pelos ensinamentos, paciência e disponibilidade. Este trabalho não teria sido possível sem seu apoio constante;

Aos colegas do Laboratório de Sementes, em especial, Felipe Sombra Cabral, que ajudou diretamente durante a instalação dos experimentos e, Liliane Santos, por estar sempre disponível em me ajudar com orientações;

Aos Laboratórios de Bioquímica e Química Analítica da UFC e UNILAB, pela realização das análises bioquímicas e espectroscopia NIR;

Aos colegas do Programa de Bioquímica da UFC, principalmente Dalton Melo, pela ajuda nas análises bioquímicas da versão anterior deste trabalho;

Ao Dr. Adenilton Camilo, pela ajuda durante a aquisição dos espectros das sementes;

Aos membros da banca examinadora, pela disponibilidade e sugestões para melhoria do trabalho;

À secretaria do PPGAF/UFC, em especial ao secretário Vitor, pela disponibilidade em sempre esclarecer e ajudar em todos os assuntos burocráticos do programa;

Aos professores do PPGAF/UFC, pelos conhecimentos adquiridos ao longo do curso;

Aos meus pais, Benedito Cartagenes e Arivanilza Costa, pelo esforço e dedicação durante meu caminho de estudos;

À minha avó, Maria José, por todo o apoio; você acreditou que eu seria uma grande mulher, muito do que sou veio de ti vó, meu grande exemplo de mulher e ser humano;

Ao meu marido, Prof. Dr. Fernando Sarmiento, pelas correções da escrita, condução dos experimentos, pelo ombro amigo diante dos choros de angústia, tristeza e dor, pela paciência, incentivo, amor e apoio incondicional;

A mim, por não desistir, parece irrisório, mas só Deus e minha família sabem da força e determinação que tive de ter para concluir este trabalho em momentos tão sombrios;

A todos aqueles que contribuíram direta ou indiretamente para a realização deste trabalho, mas não mencionados devido a um lapso de memória.

## RESUMO

Investir em técnicas mais rápidas e eficientes de avaliação do vigor de sementes garantem a alocação segura e racional de recursos e energia, promove o desenvolvimento científico e tecnológico relacionado à produção de sementes e grãos. A combinação da espectroscopia com a análise multivariada pode melhorar a avaliação e/ou seleção de sementes e grãos, minimizar o erro do analista humano e reduzir custos. Este estudo teve como objetivo, investigar o potencial da espectroscopia no infravermelho próximo, combinada com a análise multivariada, para avaliação e classificação do potencial fisiológico de lotes de sementes de gergelim e sorgo. O primeiro estudo foi realizado com nove lotes de sementes de gergelim, enquanto o segundo utilizou quatro lotes de sementes de sorgo. Os lotes foram classificados em diferentes níveis de vigor por meio da análise discriminante por regressão dos mínimos quadrados parciais (PLS-DA), usando NIR e ferramentas quimiométricas. O modelo de regressão PLS-DA, com validação cruzada, teve o melhor resultado com os pré-tratamentos segunda derivada, segunda derivada e correção de espalhamento de luz e *Smoothing* + primeira derivada para os níveis de alto, médio e baixo vigor das sementes, respectivamente, com mais de 80% de eficiência na classificação de sementes de gergelim. Para as sementes de sorgo, o resultado que mais se destacou foi com o pré-tratamento segunda derivada + SNV, com 100% de eficiência na classificação. Os principais espectros de NIR das sementes de sorgo e gergelim ocorrem nas regiões entre 997 a 1452 nm e 1040 a 1449 nm, respectivamente. A interpretação dos espectros de NIR indicam que a água foi o composto mais importante nas sementes de sorgo; para o gergelim, a presença de hidrocarbonetos alifáticos foi mais pronunciada. Os testes de validação externa indicaram que sementes de sorgo e gergelim, com diferentes níveis de vigor, foram classificadas com 100% e cerca de 80% de acertos, respectivamente. Portanto, os resultados deste estudo indicam que a combinação de espectroscopia no infravermelho próximo com a análise multivariada foi eficaz para avaliar e classificar o potencial fisiológico de lotes de sementes de gergelim e sorgo.

**Palavras-chave:** *Sesamum indicum*; *Sorghum bicolor*; PLS-DA; NIR; quimiometria.

## ABSTRACT

Investing in faster and more efficient techniques for evaluating seed vigor guarantees the safe and rational allocation of resources and energy, promotes scientific and technological development related to the production of seeds and grains. The combination of spectroscopy with multivariate analysis can improve the evaluation and/or selection of seeds and grains, minimize human analyst error and reduce costs. This study aimed to investigate the potential of near-infrared spectroscopy, combined with multivariate analysis, to evaluate and classify the physiological potential of sesame and sorghum seed lots. The first study was carried out with nine lots of sesame seeds, while the second used four lots of sorghum seeds. The lots were classified into different vigor levels using partial least squares regression (PLS-DA) discriminant analysis, using NIR and chemometric tools. The PLS-DA regression model, with cross-validation, had the best result with the pre-treatments second derivative, second derivative and light scattering correction and Smoothing + first derivative for high, medium and low seed vigor levels, respectively, with more than 80% efficiency in the classification of sesame seeds. For sorghum seeds, the result that stood out the most was with the pre-treatment second derivative + SNV, with 100% efficiency in classification. The main NIR spectra of sorghum and sesame seeds occur in the regions between 997 to 1452 nm and 1040 to 1449 nm, respectively. The interpretation of NIR spectra indicates that water was the most important compound in sorghum seeds; for sesame, the presence of aliphatic hydrocarbons was more pronounced. The external validation tests indicated that sorghum and sesame seeds, with different levels of vigor, were classified with 100% and about 80% of correct answers, respectively. Therefore, the results of this study indicate that the combination of near-infrared spectroscopy with multivariate analysis was effective to evaluate and classify the physiological potential of sesame and sorghum seed lots.

**Keywords:** *Sesamum indicum*; *Sorghum bicolor*; PLS-DA; NIR; chemometrics.

## LISTA DE FIGURAS

- Figura 1 - Espectros brutos de reflectância difusa das sementes de gergelim sem pré-tratamento, obtidos na faixa espectral de 900 a 1700 nm em espectrofotômetro portátil (NIR), para todos os níveis de vigor (A), alto vigor (B), médio vigor (C) e baixo vigor (D)..... 33
- Figura 2 - Espectros de infravermelho próximo de sementes de gergelim com diferentes níveis de vigor, com os pré-tratamentos segunda derivada Savitzky-Golay, segunda derivada Savitzky-Golay + MSC e *Smoothing* + primeira derivada Savitzky-Golay, obtidos na faixa espectral de 900 a 1700 nm em espectrofotômetro portátil (NIR), para as sementes de alto vigor (A), médio vigor (B) e baixo vigor (C), respectivamente. .... 36
- Figura 3 - Espectros de infravermelho próximo de sementes de sorgo sem pré-tratamento, obtidos na faixa espectral de 900 a 1700 nm em espectrofotômetro portátil (NIR), para todos os níveis de vigor (A) e para as amostras de alto vigor (B) e baixo vigor (C)..... 49
- Figura 4 - Espectros de infravermelho próximo de sementes de sorgo com diferentes níveis de vigor, com o pré-tratamentos com segunda derivada Savitzky-Golay + SNV, obtidos na faixa espectral de 900 a 1700nm em espectrofotômetro portátil (NIR), para todos os níveis de vigor (A), alto vigor (B) e baixo vigor (C)..... 52

## LISTA DE TABELAS

Tabela 1 - Testes de germinação (G), primeira contagem de germinação (PC), índice de velocidade de germinação (IVG), emergência de plântulas (EP), índice de velocidade de emergência (IVE), massa seca de plântulas (MS), comprimento de plântulas (CP), envelhecimento acelerado de 24 (EA24h) e 48h (EA48h). ....	31
Tabela 2 - Parâmetros de classificação obtidos do modelo PLS-DA para classificação das sementes de gergelim. ....	34
Tabela 3 - Bandas de absorção presentes nas sementes de gergelim obtidas pelo espectrofotômetro NIR. ....	37
Tabela 4 - Testes de germinação (G), primeira contagem de germinação (PC), índice de velocidade de germinação (IVG), emergência de plântulas (EP), índice de velocidade de emergência (IVE), condutividade elétrica (CE), massa seca de plântulas (MS), comprimento de plântulas (CP), envelhecimento acelerado de 24 (EA24h) e 48h (EA48h).....	47
Tabela 5 - Parâmetros de classificação obtidos do modelo PLS-DA para classificação das sementes de sorgo. ....	50

## SUMÁRIO

1	<b>INTRODUÇÃO GERAL</b> .....	15
2	<b>REFERENCIAL TEÓRICO</b> .....	16
2.1	<b>A cultura do gergelim</b> .....	16
2.2	<b>A cultura do sorgo</b> .....	17
2.3	<b>Vigor de sementes e análise de imagens</b> .....	18
2.4	<b>Espectroscopia no infravermelho próximo (NIR)</b> .....	19
2.5	<b>Quimiometria</b> .....	20
2.6	<b>PLS-DA</b> .....	21
3	<b>CAPÍTULO I: ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO E ANÁLISE MULTIVARIADA NA AVALIAÇÃO DO VIGOR DE SEMENTES DE GERGELIM</b> .....	23
4	<b>INTRODUÇÃO</b> .....	25
5	<b>MATERIAL E MÉTODOS</b> .....	26
5.1	<b>Local do estudo e material utilizado</b> .....	26
5.2	<b>Condução e delineamento experimental</b> .....	26
5.3	<b>Caracterização fisiológica dos lotes de sementes de gergelim</b> .....	27
5.3.1	<i>Determinação do teor de água (TA)</i> .....	27
5.3.2	<i>Germinação (G) e primeira contagem de germinação (PC)</i> .....	27
5.3.3	<i>Índice de velocidade de germinação (IVG)</i> .....	27
5.3.4	<i>Emergência de plântulas (EP) e Índice de velocidade de emergência (IVE)</i> .....	27
5.3.5	<i>Comprimento de plântulas (CP)</i> .....	28
5.3.6	<i>Massa seca de plântulas (MS)</i> .....	28
5.3.7	<i>Envelhecimento acelerado (EA)</i> .....	28
5.4	<b>Análise estatística para a caracterização do perfil dos lotes de sementes de gergelim</b> .....	28
5.5	<b>Obtenção dos espectros por espectroscopia NIR em sementes de gergelim</b> .....	28
5.6	<b>Análise espectral em sementes de gergelim</b> .....	29
6	<b>RESULTADOS E DISCUSSÃO</b> .....	31
6.1	<b>Caracterização dos lotes de sementes de gergelim</b> .....	31

6.2	Imagens dos espectros de absorção no infravermelho próximo em sementes de gergelim .....	32
6.3	Modelos quimiométricos de classificação dos lotes de sementes de gergelim .....	33
6.4	Interpretação dos espectros de absorção no infravermelho próximo em sementes de gergelim .....	35
7	CONCLUSÃO .....	38
8	CAPITULO II: ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO E ANÁLISE MULTIVARIADA NA AVALIAÇÃO DO VIGOR DE SEMENTES DE SORGO .....	39
9	INTRODUÇÃO .....	41
10	MATERIAL E MÉTODOS .....	42
10.1	Local do estudo e material utilizado .....	42
10.2	Condução e delineamento experimental.....	42
10.3	Caracterização fisiológica dos lotes de sementes de sorgo.....	43
10.3.1	<i>Determinação do teor de água (TA)</i> .....	43
10.3.2	<i>Germinação (G) e primeira contagem de germinação (PC)</i> .....	43
10.3.3	<i>Índice de velocidade de germinação (IVG)</i> .....	43
10.3.4	<i>Emergência (E) e Índice de velocidade de emergência (IVE)</i> .....	43
10.3.5	<i>Comprimento de plântulas (CP)</i> .....	44
10.3.6	<i>Massa seca de plântulas (MS)</i> .....	44
10.3.7	<i>Condutividade elétrica (CE)</i> .....	44
10.3.8	<i>Envelhecimento acelerado (EA)</i> .....	44
10.4	Análise estatística para a caracterização do perfil dos lotes de sementes de sorgo .....	45
10.5	Obtenção dos espectros por espectroscopia NIR em sementes de sorgo	45
10.6	Análise espectral em sementes de sorgo .....	45
11	RESULTADOS E DISCUSSÃO .....	47
11.1	Caracterização dos lotes de sementes de sorgo.....	47
11.2	Imagens dos espectros de absorção no infravermelho próximo em sementes de sorgo .....	48
11.3	Modelos quimiométricos de classificação dos lotes de sementes de sorgo.....	49

11.4	Interpretação dos espectros de absorção no infravermelho próximo em sementes de sorgo .....	51
12	CONCLUSÃO .....	53
	REFERÊNCIAS .....	54
	APÊNDICE A – RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS E DOS PRÉ-TRATAMENTOS ESPECTRAIS EM SEMENTES DE GERGELIM.....	63
	APÊNDICE B – RESULTADOS DOS PRÉ-TRATAMENTOS ESPECTRAIS EM SEMENTES DE SORGO .....	65

## 1 INTRODUÇÃO GERAL

Um dos pré-requisitos para alta produtividade de culturas agrícolas no campo é a utilização de sementes de alto vigor. Em geral, a seleção de sementes vigorosas para o plantio resulta em estandes de plantas de maior desempenho. De acordo com Fan, Ma e Wu (2020), sementes de alto vigor têm maiores chances de garantir produção de alta qualidade e boa adaptabilidade a efeitos bióticos ou abióticos, como seca, geada, doenças e pragas. No entanto, a manutenção desses altos rendimentos requer o estabelecimento bem-sucedido e uniforme de plantas no campo sob condições ambientais variadas (REED; BRADFORD; KHANDAY, 2022).

O vigor das sementes é uma variável complexa relacionada à longevidade das sementes, velocidade de germinação, crescimento das plântulas e tolerância ao estresse precoce (REED; BRADFORD; KHANDAY, 2022). A perda de vigor pela semente pode ser devido a uma série de razões, incluindo o superaquecimento durante a secagem, danos físicos durante o processamento pós-colheita e envelhecimento durante o armazenamento (AMBROSE *et al.*, 2016). A deterioração das sementes durante o armazenamento modifica as membranas celulares e conseqüentemente reduz seu vigor. Outros processos também são encontrados, como mutações, degradação de proteínas e inativação de enzimas (RATAJCZAK *et al.*, 2015; BARRETO e GARCIA, 2017).

Existem diversos testes para avaliar a viabilidade e o potencial fisiológico de lotes de sementes, como os testes de germinação e os testes de vigor, incluindo envelhecimento acelerado, condutividade elétrica, emergência de plântulas, primeira contagem de germinação, índice de velocidade de germinação e massa seca de plântulas. Esses testes podem classificar efetivamente os lotes com base em seus níveis de vigor, no entanto, são demorados, destrutivos e onerosos em materiais e indivíduos, e estão sujeitos à subjetividade da avaliação.

Com base nesse pressuposto, para melhorar a rapidez e a eficiência dos sistemas de avaliação da qualidade de sementes, novos estudos são incentivados para determinar o potencial de técnicas mais rápidas e eficientes de avaliação não destrutiva do vigor de sementes para garantir a alocação segura e o uso racional de recursos e energia, e promover o desenvolvimento científico e tecnológico relacionado à produção de sementes. Neste sentido, a combinação da espectroscopia com a análise multivariada pode melhorar a avaliação e/ou seleção de sementes e minimizar o erro do analista humano.

Estudos anteriores indicam que o uso da espectroscopia no infravermelho próximo (NIR) mostrou eficiência em testes de viabilidade e classificação de sementes, como demonstrado em sementes de milho, no qual a técnica foi aplicada em sementes de diferentes

colorações: branco, amarelo e roxo, e os resultados da análise indicaram que a espectroscopia FT-NIR classificou corretamente as sementes viáveis e não viáveis para todas as três categorias, com precisão de 100% e capacidade preditiva de mais de 95%, utilizando a análise multivariada de mínimos quadrados parciais (AMBROSE *et al.*, 2016). Ao utilizar dados de FT-NIR, Medeiros *et al.* (2020) verificaram que a precisão da predição de germinação e vigor em sementes de *Urochloa brizantha* atingiu 82 e 61%, respectivamente. Para sementes de trigo, a espectroscopia NIR previu o vigor das sementes em três diferentes níveis (FAN, MA e WU, 2020).

O NIR permite prever certas substâncias, qualitativa ou quantitativamente, baseando-se no princípio de comparação dos resultados obtidos em laboratório com os espectros obtidos pelo equipamento (VIANA *et al.*, 2010). A tecnologia NIR é baseada nas leituras obtidas na região espectral do infravermelho próximo (700-2500 nm), em que a radiação infravermelha penetra na amostra e interage com as ligações constituintes. Essa interação tem características específicas de oscilação e frequência, de modo que pode ser medida no modo de transmitância/refletância e, em seguida, correlacionada com os grupos funcionais (como ligações C-H, S-H, N-H, e O-H) na molécula, e os resultados podem ser analisados usando análise multivariada (PASQUINI, 2003; GUIMARÃES, 2016).

O uso do NIR representa a capacidade de realizar análises continuamente em um curto período de tempo, o que pode gerar grande quantidade de informações, reduzindo a necessidade de mão-de-obra. Além disso, não polui o meio ambiente e não utiliza produtos químicos ou reagentes (MAYRINCK *et al.*, 2020).

Portanto, o objetivo desse estudo foi investigar o potencial da espectroscopia no infravermelho próximo (NIR), combinada com a análise multivariada, para a avaliação e classificação do potencial fisiológico de lotes de sementes de gergelim e sorgo.

## **2 REFERENCIAL TEÓRICO**

### **2.1 A cultura do gergelim**

O gergelim (*Sesamum indicum* L.) é uma espécie oleaginosa, pertencente à família Pedaliaceae, originário da Ásia e da África, de onde se disseminou para todos os países de clima quente e no Brasil, foi introduzido pelos portugueses no século XVI (ARRIEL; BELTRÃO; FIRMINO, 2009). A produção mundial de gergelim é de 5,5 mil toneladas, em uma área cultivada de 9,9 milhões de hectares, com produtividade média de 554,1 kg ha<sup>-1</sup> (FAOSTAT, 2019). No Brasil, a produção de gergelim na safra 2020/2021 foi de 98,1 mil toneladas em uma

área plantada de 149,9 mil hectares, com destaque para os estados de Mato Grosso, Mato Grosso do Sul e Goiás (CONAB, 2022).

A espécie é conhecida por sua tolerância ao estresse hídrico, facilidade de cultivo e produtividade, pois sob condições de chuvas de 300 mm, desde que bem distribuída ao longo de seu desenvolvimento, a espécie ainda mantém altos rendimentos (SILVA *et al.*, 2016). Portanto, o gergelim é usado como cultura alternativa para exploração em pequenas comunidades rurais, representando um importante centro econômico de apoio no semiárido nordestino (SILVA *et al.*, 2017; CRUZ *et al.*, 2013). Por outro lado, o gergelim não é tolerante a deficiência de oxigênio no solo, chuvas intensas ou irrigação inadequada podem levar ao encharcamento do solo e conseqüentemente diminuição da sua oxigenação, tornando-se um dos principais fatores limitantes para a cultura (CRUZ *et al.*, 2019).

Suas sementes possuem alto teor de óleo e excelente qualidade, tendo sido utilizadas para consumo humano e outras finalidades, como industrial, farmacêutico, medicinal e, mais recentemente, para a produção de energia sustentável por meio de biocombustíveis (CRUZ *et al.*, 2019).

## 2.2 A cultura do sorgo

*Sorghum bicolor* (L.) Moench é um importante cereal para as regiões tropicais e subtropicais do mundo. Semelhante ao milho, seu valor nutritivo está entre 90 e 95%, sendo considerado equivalente em minerais e vitaminas (MOREIRA *et al.*, 2014), um substituto deste cereal em regiões de climas semiáridos. Nestas áreas, devido à maior resistência a seca e maior rendimento por unidade de área, o sorgo tem sido mais utilizado (SOUSA *et al.*, 2018). Por se originar em um ambiente árido, e por ser uma planta C<sub>4</sub>, é uma espécie que pode tolerar condições de baixa disponibilidade de água, por isso tem a capacidade de crescer sob altas temperaturas e baixa pluviosidade ou em condições irregulares de chuva (SARMENTO *et al.*, 2020).

No Brasil, a cultura do sorgo ocupava uma área de 1.060,1 mil hectares na safra 2020/2021, produzindo aproximadamente 2.924,8 milhões de toneladas de grãos, com produtividade média de 2.759 kg ha<sup>-1</sup> (CONAB, 2022). A cultura tem sido implementada no sistema de integração lavoura-pecuária, juntamente com a braquiária para a produção de grãos e forragem para ser usada como pastejo, melhorando a disponibilidade de alimentos na entressafra, reduzindo assim a flutuação na oferta de animais para abate ao longo do ano, e ainda contribui para a conservação do solo, com os restos das palhas que servem para manutenção do sistema de plantio direto (SILVA *et al.*, 2014). Além disso, se caracteriza como

opção para cultivo no período da safrinha por apresentar características xerófitas, a qual confere tolerância à ambiente com déficit hídrico, comuns nos plantios de segunda safra na região sudoeste de Goiás (MARTINS *et al.*, 2016).

A composição química e nutricional do sorgo é rica em nutrientes e compostos bioativos, que podem variar conforme os genótipos, condições edafoclimáticas e práticas de adubação. Por exemplo, Kardeş *et al.* (2021), ao avaliar 156 genótipos de sorgo, relataram teores de óleo bruto de grãos de sorgo próximo de 4%; teores de proteína bruta de aproximadamente 9,6%; teor de amido próximo de 77%.

### 2.3 Vigor de sementes e análise de imagens

Os testes de vigor auxiliam os resultados do teste de germinação, pois detectam com maior precisão os avanços da deterioração e avaliam o potencial fisiológico das sementes em lotes com porcentagem de germinação similar. Lotes de sementes vigorosas são capazes de se estabelecer em ambientes com condições desfavoráveis para sua espécie de forma rápida e eficaz. Dentre esses aspectos, tem-se desenvolvido e adaptado diversos testes para avaliação do vigor de lotes de sementes, sendo considerados eficientes os que permitem a separação de lotes em diferentes categorias de vigor, desde que essas informações correspondam aos resultados dos testes tradicionais (MARCOS- FILHO, 2016).

Alguns testes como condutividade elétrica (SOUSA *et al.*, 2018), tetrazólio (CARVALHO *et al.*, 2018), envelhecimento acelerado (LIMA *et al.*, 2019), lixiviação de potássio (MEDEIROS *et al.*, 2019), pH do exsudato (ALVES *et al.*, 2016), deterioração controlada (LEÃO-ARAÚJO *et al.*, 2017) são testes de vigor que se destacam para caracterização de lotes de sementes. No entanto, estes testes são demorados, e o cenário produtivo requer, cada vez mais, resultados rápidos e eficientes para a tomada de decisão no campo.

Dessa forma, o uso de técnicas de análise de imagens vêm crescendo no campo de análise de sementes, trazendo tecnologias não destrutivas e rápidas, por meio da avaliação de imagens de sementes e plântulas, para a detecção da viabilidade e diferenças de vigor entre lotes de sementes de diversas espécies, como o teste de Raios-X, que possibilita observar as estruturas internas das sementes e examinar seu preenchimento anatomicamente e características morfológicas (MEDEIROS *et al.*, 2020); imagem multiespectral e hiperespectral, serve para o desenvolvimento de sistema industrial, em grande escala construído sobre imagem hiperespectral para classificação rápida e não destrutiva de uma ampla gama de produtos de acordo com seus fatores internos (MUKASA *et al.*, 2020; WAKHOLI *et al.*, 2018);

e o uso do infravermelho próximo que, por meio da penetração da radiação infravermelha na amostra, permite relacionar os constituintes orgânicos para a avaliação (GRUNVALD *et al.*, 2014).

Além das técnicas usadas diretamente nas sementes, a análise de imagens computadorizadas de plântulas, por meio de *softwares*, como o *Seed Vigor System (SVIS®)* (BARBOSA *et al.*, 2016; LEÃO-ARAÚJO *et al.*, 2017), *Image Processing and Analysis In Java (Image®J)* (SILVA e DOTTO, 2017), Sistema de Análise de Plântulas (SAPL®) (MEDEIROS e PEREIRA, 2018); Sistema Automatizado do Vigor de Sementes (VigorS®) (CASTAN; GOMES-JUNIOR; MARCOS-FILHO, 2018) e *GroundEye®* (PINTO *et al.*, 2015) permitem por meio da avaliação do desenvolvimento de plântulas, inferir o vigor de lotes de sementes, analisar um volume alto de amostras de forma rápida e precisa, e fornecer um banco de imagens da espécie estudada, sendo, portanto, técnicas eficientes na avaliação do vigor de sementes.

#### **2.4 Espectroscopia no infravermelho próximo (NIR)**

A espectroscopia é um termo atribuído pela Física e pela Química para todas as técnicas de obtenção de dados através da radiação incidente em uma amostra, baseando-se na aplicação da quimiometria. O espectro eletromagnético é o intervalo de todas as frequências de ondas eletromagnéticas existentes desde as ondas de rádio a radiação gama. A região do infravermelho é compreendida entre os comprimentos de onda de 700 nm a  $1 \times 10^6$  nm, e é dividida em: infravermelho próximo (*Near infrared* – NIR), que corresponde ao comprimento de onda de 700 a 2500 nm; infravermelho médio (*Middle infrared* – MID), de 2500 a 10000 nm; e infravermelho distante (*Far infrared* – FIR) de 10000 a  $1 \times 10^6$  nm (PASQUINI, 2003; REICH 2005).

A radiação incidente na amostra promove a interação da radiação eletromagnética com materiais constituídos por ligações hidrogeniônicas, como Carbono-Hidrogênio (C-H), Oxigênio-Hidrogênio (O-H), Enxofre-Hidrogênio (S-H) e Nitrogênio-Hidrogênio (N-H), que permitem obter informações moleculares (PASQUINI, 2003; GUIMARÃES, 2016; KUMAR *et al.*, 2016). Por ser um método de análise rápido, de alta precisão, não destrutivo e de baixo custo, a espectroscopia no infravermelho próximo tem sido eficiente em estudos de produtos agrícolas e solos, pois seu uso tem grande influência sobre material composto por matéria orgânica (ALVAREZ, 2018).

Em plantas e, principalmente, em forragem, o NIR é normalmente utilizado para a análise de nitrogênio, umidade, fibra, carboidratos estruturais, aminoácidos e minerais para

monitorar a qualidade (QUENTIN *et al.*, 2016). Em frutos, tem sido empregado para prever os compostos do metabolismo primário (firmeza, pH, açúcar solúvel total, sólido solúvel (SS), acidez titulável (AT), relação SS/AT), e secundário (vitamina C, carotenoides totais, flavonoides totais, polifenóis totais e atividade antioxidante total), como foi visto por Samamad *et al.* (2018), em estudo de fisiologia e fenotipagem de clones precoces de cajueiro anão. Nas sementes, as informações moleculares são interpretadas por meio dos elementos químicos e ondas vibracionais, que podem explorar diferentes composições bioquímicas existentes entre as sementes novas e envelhecidas, como os teores de amido, proteína e umidade (AMBROSE *et al.*, 2016).

Os primeiros trabalhos que utilizaram essa técnica em sementes, foram realizados para determinação do teor de água de sementes de soja e milho por Hart, Norris e Golubic, em 1962. Com o avanço da utilização dessa técnica, é possível aplicá-la em estudos sobre viabilidade e qualidade fisiológica de sementes, conforme observado por Alvarez (2018), que estudou a qualidade fisiológica de sementes de milho e soja por espectroscopia no infravermelho próximo e, Olesen *et al.* (2011), ao avaliarem a viabilidade de sementes de espinafre.

## 2.5 Quimiometria

A quimiometria é uma disciplina da química que emprega métodos matemáticos e estatísticos para planejar ou selecionar experimentos de forma otimizada e para fornecer o máximo de informação química com a análise dos dados obtidos (FERREIRA *et al.*, 1999). Devido as informações extraídas do NIR não poderem ser diretamente analisadas, em função da complexidade dos dados gerados e às inúmeras interferências que ocorrem, é necessário pré-processar e corrigir o espectro para avaliação.

A correção dos espectros consiste na utilização de técnicas que permitem reduzir, eliminar ou padronizar o impacto desses fatores não desejáveis nos espectros, sem alterar a informação espectroscópica contida neles (SIESLER *et al.*, 2002). São exemplos de pré-tratamentos: filtro de média móvel, derivada de Savitsky-Golay (MALINOWSKI, 1991), alisamento e correção do espalhamento multiplicativo do sinal (*Multiplicative Scatter Correction*- MSC) e variação de padrão normal (*Standard Normal Variate* – SNV) (PANTOJA, 2006). Usando o espectro corrigido, os dados podem ser analisados. Para isso, é necessário estabelecer um modelo de calibração. Métodos multivariados são os mais apropriados porque permitem o estudo de várias espécies ao mesmo tempo, independentemente de haver diferenças

espectrais significativas entre eles ou se há um alto grau de correlação nos dados (MAYRINCK *et al.*, 2020).

A calibração requer entrada de trabalho inicial, mas uma vez calibrada, o sistema NIR é eficiente, rápido e seguro (KAUR; SANGHA; KAUR, 2017). De acordo com Burns e Ciurczak (2008), as principais técnicas de análise multivariada utilizada para a elaboração de modelos são: regressão linear múltipla (*Multiple Linear Regression - MLR*), regressão por componentes principais (*Principal Components Regression - PCR*), análise por componentes principais (*Principal Component Analysis - PCA*) e mínimos quadrados parciais (*Partial Least Square - PLS*). Ainda podem ser utilizados vários algoritmos, como análise discriminante linear (*Linear Discriminant Analysis - LDA*), floresta aleatória (*Random Forest - RF*), Bayes ingênuo (*Naive Bayes - NB*), máquina de vetores de suporte com kernel linear (*Support Vector Machine - SVM-l*) e base radial (*SVM-r*), e rede neural artificial (RNA) (MEDEIROS *et al.*, 2020).

Após a etapa de calibração, é necessário avaliar o modelo a partir da validação interna e externa para indicar se ele pode determinar os parâmetros necessários. Para verificação interna, em um conjunto de  $n$  amostras, uma das amostras é sempre reservada como não participante da calibração (*leave one out*), portanto, as propriedades da amostra são previstas. Para validação externa, utilizam-se amostras que não participaram da construção do modelo de calibração (BERRUETA *et al.*, 2007; SILVA, 2019).

## 2.6 PLS-DA

A Análise Discriminante de Mínimos Quadrados Parciais (PLS-DA) é um método supervisionado desenvolvido a partir do mesmo algoritmo PLS (*Partial Least Squares - Mínimos Quadrados Parciais*) usado para calibração multivariada e comumente aplicado para reduzir a dimensionalidade de um banco de dados, podendo ser adaptado para a finalidade de seleção e classificação de amostras (RUIZ-PEREZ *et al.*, 2020).

Quando usado para classificação, o modelo PLS-DA determina a qual classe uma amostra desconhecida pertence com base nas informações fornecidas ao sistema (SANTANA *et al.*, 2020). Ao contrário do PLS, a variável de resposta  $Y$  no PLS-DA é um conjunto de variáveis fictícias que representam categorias de amostras, geralmente atribuindo o número 1 à categoria de interesse e 0 às demais categorias. O valor previsto no PLS-DA é um número real, não um número inteiro imaginário.

Portanto, é necessário definir um valor de corte para determinar a qual classe uma amostra pertence, geralmente definir o valor de corte para 0,5, e usar o método de validação cruzada completa para calcular o PLS-DA (KONG *et al.*, 2013).

Dois conjuntos de métodos, calibração e validação, são elaborados no método PLS-DA. O conjunto de calibração é usado para desenvolver um modelo de classificação e encontrar os melhores parâmetros para classificação. O conjunto de validação consiste em amostras não utilizadas no conjunto de calibração que deve ser fornecida para estimar o desempenho da equação de classificação avaliada em amostras futuras. A precisão de calibração e validação do modelo é calculada com base no número de amostras identificadas corretamente usando vários métodos de pré-processamento. Uma redução completa dos dados é então realizada para criar variáveis latentes que se correlacionam entre si e descrevem a variável resposta (KUSUMANINGRUM *et al.*, 2017).

### **3 CAPÍTULO I: ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO E ANÁLISE MULTIVARIADA NA AVALIAÇÃO DO VIGOR DE SEMENTES DE GERGELIM**

#### **RESUMO**

A utilização de novas metodologias em tecnologia de sementes, como o uso de espectroscopia no infravermelho próximo (NIR) em conjunto à quimiometria, pode melhorar a velocidade, eficiência e automação do processo de avaliação do controle de qualidade de sementes. Este estudo objetivou avaliar a eficiência da combinação da técnica NIR e a análise multivariada na avaliação do vigor de sementes de gergelim. Para tanto, foram utilizados nove lotes de semente de gergelim, que apresentavam diferentes níveis de qualidade fisiológica. Para avaliar o potencial fisiológico inicial das sementes foram realizados testes de germinação, primeira contagem de germinação, índice de velocidade de germinação, emergência de plântulas, índice de velocidade de emergência, massa seca de plântulas, comprimento de plântulas e envelhecimento acelerado. Os lotes foram classificados em diferentes níveis de vigor por meio da análise discriminante por regressão dos mínimos quadrados parciais (PLS-DA), empregando espectroscopia no infravermelho próximo e ferramentas quimiométricas. O modelo de regressão PLS-DA com validação cruzada teve o melhor resultado com os pré-tratamentos segunda derivada, segunda derivada e correção de espalhamento de luz e *Smoothing* + primeira derivada para os níveis de alto, médio e baixo vigor das sementes, respectivamente, com mais de 80% de eficiência na classificação de lotes de sementes de gergelim de diferentes níveis de vigor. Esses resultados indicam que a combinação da técnica NIR e a análise multivariada é eficiente para avaliar o vigor de lotes de sementes de gergelim, além de rápida e não destrutiva.

**Palavras-chave:** *Sesamum indicum*; NIR; quimiometria; PLS-DA.

## CHAPTER I: NEAR INFRARED SPECTROSCOPY AND MULTIVARIATE ANALYSIS IN ASSESSING THE VIGOR OF SESAME SEEDS

### ABSTRACT

The use of new methodologies in seed technology, such as the use of near-infrared spectroscopy (NIR) in conjunction with chemometrics, can improve the speed, efficiency and automation of the seed quality control evaluation process. This study aimed to evaluate the efficiency of the combination of the NIR technique and the multivariate analysis in the evaluation of the vigor of sesame seeds. For that, nine lots of sesame seeds were used, which had different levels of physiological quality. To evaluate the initial physiological potential of the seeds, germination tests, first germination count, germination speed index, seedling emergence, emergence speed index, seedling dry mass, seedling length and accelerated aging were performed. The lots were classified into different vigor levels by means of discriminant analysis by partial least squares regression (PLS-DA), using near-infrared spectroscopy and chemometric tools. The PLS-DA regression model with cross-validation had the best result with the pre-treatments second derivative, second derivative and light scattering correction and Smoothing + first derivative for high, medium and low seed vigor levels, respectively, with more than 80% efficiency in classifying lots of sesame seeds of different vigor levels. These results indicate that the combination of the NIR technique and the multivariate analysis is efficient to evaluate the vigor of sesame seed lots, in addition to being fast and non-destructive.

**Key words:** *Sesamum indicum*; NIR; chemometrics; PLS-DA.

## 4 INTRODUÇÃO

O gergelim (*Sesamum indicum* L) é uma das principais oleaginosas cultivadas no mundo e uma das primeiras espécies domesticadas (SHARABY; BUTOVCHENKO, 2019). Suas sementes possuem alto valor nutricional (fonte de cálcio, fósforo e proteína) e são utilizadas em todo o mundo para diversas finalidades, incluindo alimentação humana, na forma crua ou torrada, no preparo de alimentos, molhos e saladas; e propriedades medicinais (FREITAS *et al.*, 2018). As sementes também são utilizadas para produzir óleo de alta qualidade, até superior a outras oleaginosas (MYINT *et al.*, 2020).

O aumento da demanda por gergelim nos últimos anos teve impacto positivo nos volumes produzidos e nas áreas plantadas. As plantações de gergelim no Brasil aumentaram de 53 mil hectares em 2018/2019 para 175 mil hectares em 2019/2020, um aumento de 230%, enquanto a produção aumentou 123%, saindo de 41 mil para 95,8 mil toneladas de grãos (CONAB, 2021). No entanto, estes números de crescimento animadores não escondem o fato de que o manejo da cultura e a produção de sementes de alta qualidade ainda não são totalmente elucidados, necessitando de conhecimento técnico e investimentos em novas tecnologias, por exemplo, o desenvolvimento de testes para avaliar de maneira rápida e eficiente o potencial fisiológico das sementes (XAVIER *et al.*, 2019; LARIOS *et al.*, 2020b).

Estudos recentes demonstraram a utilização de técnicas para ajudar a classificar, visualizar e padronizar as avaliações em testes físicos e fisiológicos de sementes, como análise de imagens para a sua classificação biométrica, o que pode reduzir a subjetividade das avaliações visuais (XAVIER *et al.*, 2019); e o uso da espectroscopia de infravermelho próximo (NIR) para determinar os grupos funcionais da constituição química de amostras orgânicas e inorgânicas (VASCONCELOS *et al.*, 2018), identificando sementes com diferentes níveis de vigor (LARIOS *et al.*, 2020a).

O NIR tem se mostrado eficaz na substituição de métodos qualitativos e quantitativos de análise laboratoriais em diversos setores, incluindo análises de alimentos, produtos farmacêuticos e agrícolas, permitindo a avaliação de diversas propriedades e características em um amplo espectro de amostras (KUSUMANINGRUM *et al.*, 2017; PASQUINI, 2018; SIMEONE *et al.*, 2019; MEENU *et al.*, 2022).

Em sementes, a utilização do NIR associada a métodos de calibração multivariada tem sido aplicada para distinguir características químicas em amarantos (XAVIER *et al.*, 2019), diferenciar cultivares de girassol, tanto por meio de óleo, quanto por sementes (VASCONCELOS *et al.*, 2018), analisar a qualidade de sementes de algodão em função de diferentes níveis de vigor (MAYRINCK *et al.*, 2020), discriminar sementes de soja com alto e

baixo vigor (LARIOS *et al.*, 2020) e identificar padrão de qualidade de café agroflorestal especial (MANUEL *et al.*, 2022). No entanto, o emprego dessa técnica em estudos de sementes de gergelim ainda não foi explorado.

A utilização do NIR justifica-se por ser um procedimento rápido para obtenção de resultados, não ser destrutivo e não necessitar de reagentes que geram resíduos, além de poder utilizar diversas características constitutivas qualitativas e quantitativas ao mesmo tempo, bem como obter espectros em amostras sólidas e líquidas sem necessitar de maiores preparos (VASCONCELOS *et al.*, 2018; SIMEONE *et al.*, 2019; MAYRINCK *et al.*, 2020). A hipótese é que acreditamos que o NIR pode ser utilizado como alternativa aos testes fisiológicos tradicionais, como condutividade elétrica (SOUSA *et al.*, 2018), envelhecimento acelerado (LIMA *et al.*, 2019) e tetrazólio (CARVALHO *et al.*, 2018), nos quais há perda de material vegetal e necessidade maior de tempo para avaliação.

Portanto, este estudo objetivou avaliar a eficiência da combinação da técnica NIR e a análise multivariada na avaliação do vigor de sementes de gergelim.

## **5 MATERIAL E MÉTODOS**

### **5.1 Local do estudo e material utilizado**

O estudo foi realizado nos Laboratórios de Análise de Sementes do Departamento de Fitotecnia do Centro de Ciências Agrárias da Universidade Federal do Ceará (UFC), onde foi realizada a caracterização fisiológica das sementes de gergelim, e no Laboratório de Química Analítica da Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB), no qual foi realizada a obtenção dos dados espectrais no infravermelho próximo. Foram utilizados nove lotes de sementes de gergelim (*Sesamum indicum* L.), gentilmente cedidos pela da Universidade Federal Rural do Semi-Árido (UFERSA).

### **5.2 Condução e delineamento experimental**

Inicialmente, foram caracterizados e diferenciados os perfis dos lotes, por meio da determinação do teor de água, teste de germinação e dos testes de vigor (primeira contagem de germinação, índice de velocidade de germinação, emergência de plântulas, índice de velocidade de emergência, comprimento de plântulas, massa seca de plântulas e envelhecimento acelerado).

O delineamento experimental utilizado para a caracterização dos lotes foi o inteiramente casualizado, com quatro repetições de 50 sementes.

Após a caracterização inicial dos lotes, os dados espectrais na região do NIR foram analisados para confecção dos modelos de classificação.

### **5.3 Caracterização fisiológica dos lotes de sementes de gergelim**

#### **5.3.1 Determinação do teor de água (TA)**

O teor de água foi determinado pelo método da estufa a  $105 \pm 3$  °C por 24h, com duas subamostras de 5 g de sementes por lote, antes e após o envelhecimento acelerado. Os resultados foram representados em porcentagem do peso úmido (BRASIL, 2009).

#### **5.3.2 Germinação (G) e primeira contagem de germinação (PC)**

As sementes foram higienizadas com hipoclorito de sódio a 2% por três minutos antes de serem enxaguadas em água destilada, secas com papel toalha e mantidas em ambiente controlado até o início do experimento. As sementes foram semeadas em rolo de papel umedecido com água destilada a uma taxa de 2,5 vezes o peso do papel seco e armazenadas em câmaras de germinação do tipo BOD (*Biochemical Oxygen Demand*) a 25 °C sob luz contínua (NERY *et al.*, 2018), com a contagem inicial de germinação no terceiro dia e a contagem final no sexto dia após a instalação do teste. Os resultados foram expressos em porcentagem de plântulas normais em cada lote (BRASIL, 2009).

#### **5.3.3 Índice de velocidade de germinação (IVG)**

Foi realizado concomitantemente ao teste de germinação, com contagens diárias de plântulas normais até o sexto dia após a instalação do teste. O índice foi determinado somando-se as razões do número de plântulas germinadas diariamente ao número de dias em que a germinação ocorreu (MAGUIRE, 1962).

#### **5.3.4 Emergência de plântulas (EP) e Índice de velocidade de emergência (IVE)**

As sementes foram semeadas em canteiro de areia de 1,0 m de largura por 10 m de comprimento a uma profundidade de 2,0 cm. A rega manual foi usada para regular a umidade do substrato. Foram realizadas avaliações diárias, calculando-se o número de plântulas que apresentavam os cotilédones acima do solo até o sexto dia após a semeadura. O resultado foi calculado em porcentagem de acordo com a fórmula publicada por Maguire (1962).

### 5.3.5 Comprimento de plântulas (CP)

Ao término do teste de germinação, o comprimento das plântulas foi analisado por meio de análise de imagens. Em cada lote, foram selecionadas ao acaso 10 plântulas normais. As plântulas de gergelim foram transferidas para papel A3 azul para captura digital das imagens, utilizando um *scanner*, modelo HP *Scanjet* 2004, posicionado invertido dentro de uma caixa de alumínio. As imagens capturadas foram armazenadas no formato JPEG com resolução digital de 300 dpi e examinadas manualmente com o *software* ImageJ® (*Image Processing and Analysis in Java*) (FERREIRA; RASBAND, 2012). Os dados obtidos foram representados em cm plântula<sup>-1</sup>.

### 5.3.6 Massa seca de plântulas (MS)

Após a medição do comprimento das plântulas, elas foram acondicionadas em sacos de papel e colocadas em estufa de circulação forçada de ar a  $65 \pm 3$  °C por 72 horas. A massa seca foi determinada em balança analítica de precisão (0,0001 g), e os resultados foram representados em g plântula<sup>-1</sup> (KRZYZANOWSKI *et al.*, 2020).

### 5.3.7 Envelhecimento acelerado (EA)

As sementes de gergelim foram submetidas ao envelhecimento acelerado pelo método tradicional, com umidade relativa em torno de 100%. As sementes foram colocadas em uma tela de alumínio sobrepostas em caixas plásticas (11 × 11 × 3,5 cm) tipo *gerbox*®, contendo 40 mL de água destilada no fundo, e acondicionadas em câmara de germinação tipo BOD a 41 °C por 24 e 48 horas.

## 5.4 Análise estatística para a caracterização do perfil dos lotes de sementes de gergelim

Os resultados obtidos foram submetidos à análise de variância pelo teste F ( $p \leq 0,05$ ). Em caso de significância, os valores médios dos componentes qualitativos foram comparados pelo teste *t* ( $p \leq 0,05$ ), utilizando o *software* estatístico SISVAR® (FERREIRA, 2011).

## 5.5 Obtenção dos espectros por espectroscopia NIR em sementes de gergelim

Os espectros na região do infravermelho próximo foram obtidos após homogeneização dos lotes, a partir de amostras de 1 g de sementes inteiras de gergelim para cada lote, em ambiente à temperatura de 20 °C, utilizando o espectro NIR/DLP® (NIRscan™

Nano Evaluation Module Texas Instruments), na faixa de 900 a 1700 nm, com incremento/largura de 7 nm, 50 varreduras e resolução digital de 228 comprimentos de onda.

Para a leitura dos espectros, as amostras foram colocadas em Becker de 5 mL, tendo como espectro de referência (branco) o Teflon. Os dados espectrais foram obtidos em triplicatas, e as médias dos espectros totalizaram ao final 27 amostras espectrais, que foram utilizadas para realizar os pré-tratamentos dos dados, análise do comportamento espectral e criação de modelos de classificação, utilizando o pacote do *software The Unscrambler*<sup>®</sup> X versão 10.4 (Camo Analytics, Oslo, Norway).

### 5.6 Análise espectral em sementes de gergelim

Para encontrar o melhor modelo de classificação para cada nível de vigor, diferentes pré-tratamentos foram utilizados nos dados espectrais para corrigir efeitos de espalhamento de luz e as demais influências físicas que não estejam relacionadas com a composição química das sementes, como correção multiplicativa de espalhamento (*Multiplicative Scatter Correction – MSC*), variação de padrão normal (*Standard Normal Variate – SNV*), primeira e segunda derivada Savtizzy-Golay com polinômio de segunda ordem e janela de 11 pontos. Após a aplicação dos pré-tratamentos, as amostras foram classificadas de acordo com o nível de vigor para cada pré-tratamento dos espectros utilizando a técnica de classificação supervisionada e o algoritmo Análise Discriminante por Mínimos Quadrados Parciais (*Partial Least Squares - Discriminant Analysis, PLS-DA*).

O conjunto das amostras foi dividido em dois conjuntos para construção dos modelos de classificação, sendo 70% das amostras de sementes utilizadas para treinamento (calibração), com validação interna cruzada e 30% das amostras utilizadas para validação externa. Desta forma, os modelos foram estudados, sendo os níveis de vigor as classes do vetor Y (variáveis dependentes) e os espectros obtidos constituindo o vetor X (variáveis independentes) (ABDI, 2003).

Em relação à modelagem PLS-DA, três modelos de classificação para cada nível de vigor (alto, médio e baixo vigor) foram obtidos; para a modelagem do nível de alto vigor, o valor 1 foi atribuído as amostras de alto vigor e o valor -1 para as demais amostras; já para a modelagem do nível de médio vigor, o valor 1 foi atribuído as amostras de médio vigor e o valor -1 para as demais amostras; e para a modelagem do nível de baixo vigor, o valor 1 foi atribuído as amostras de baixo vigor e o valor -1 para as demais amostras. O conjunto de treinamento foi utilizado para obter e escolher o número de variáveis latentes – PLS (VLs), enquanto o grupo de validação externa foi utilizado para validar o poder preditivo dos modelos. A escolha do melhor

pré-tratamento foi realizada com base no menor erro do modelo e o maior número de acertos de validação.

A Matriz de Confusão foi criada conforme procedimentos descritos por Botelho *et al.* (2015) para avaliar o desempenho e eficiência do modelo de classificação, por meio dos seguintes parâmetros: taxas de falso positivo (TFP), falso negativo (TFN), taxa de sensibilidade (SEN), taxa de especificidade (ESPEC), a taxa de eficiência (TEF) em %.

A taxa de falso positivo é a porcentagem de amostras negativas que foram classificadas como amostras positivas, e é calculada como a relação entre o número absoluto de falsos positivos (FP) e a soma do número absoluto de erros falsos positivos (FP) e verdadeiros negativos (VN) multiplicada por 100, representada pela equação 1:

$$\mathbf{TFP = FP/(FP+VN) \times 100} \quad \mathbf{(1)}$$

Da mesma forma, a taxa de falso negativo é a porcentagem de amostras positivas que foram classificadas como amostras negativas, sendo calculada como a relação entre o número absoluto de falsos negativos (FN) e a soma do número absoluto de erros falsos negativos (FN) e verdadeiros positivos (VP) multiplicada por 100, representada pela equação 2:

$$\mathbf{TFN = FN/(FN+VP) \times 100} \quad \mathbf{(2)}$$

A sensibilidade é a porcentagem de amostras positivas que foram classificadas corretamente como sendo dessa classe, sendo calculada como a razão entre o número absoluto de verdadeiros positivos (VP) e a soma do número absoluto de verdadeiros positivos (VP) e dos erros falsos negativos (FN) multiplicada por 100, representada pela equação 3:

$$\mathbf{SEN=VP/(VP+FN) \times 100} \quad \mathbf{(3)}$$

A especificidade é a porcentagem de amostras negativas que foram classificadas corretamente pelo modelo como as amostras negativas, sendo calculada pela razão entre o número absoluto de verdadeiros negativos (VN) e a soma do número absoluto de verdadeiros negativos (VN) e dos erros falso positivos (FP) multiplicado por 100, representada pela equação 4:

$$\mathbf{ESPEC=VN/(VN+FP) \times 100} \quad \mathbf{(4)}$$

A eficiência combina todas as informações transportadas pelas taxas de falso positivo e negativo, sendo calculada pela diferença entre o valor de 100% e a soma das taxas de erros falsos negativos (TFN) e falsos positivos (TFP), representada pela equação 5:

$$\text{TEF}=100-(\text{TFN}+\text{TFP}) \quad (5)$$

Os pré-tratamentos e os modelos de classificação dos dados foram realizados por *software The Unscrambler*<sup>®</sup> versão 10.4X (Camo Analytics, Oslo, Norway).

## 6 RESULTADOS E DISCUSSÃO

### 6.1 Caracterização dos lotes de sementes de gergelim

Os valores dos parâmetros de verificação do vigor das sementes e o teste estatístico de hipótese de semelhança encontram-se na Tabela 1. Todas as variáveis apresentaram diferenças significativas em todos os lotes. Esses resultados indicam que a avaliação do vigor das sementes por métodos clássicos nos fornece informações sensíveis sobre os níveis de deterioração e viabilidade, além de auxiliar na seleção de sementes vigorosas para o cultivo.

Tabela 1 - Testes de germinação (G), primeira contagem de germinação (PC), índice de velocidade de germinação (IVG), emergência de plântulas (EP), índice de velocidade de emergência (IVE), massa seca de plântulas (MS), comprimento de plântulas (CP), envelhecimento acelerado de 24 (EA24h) e 48h (EA48h).

LOTES	VIGOR	G	PC	IVG	EP	IVE	MS	CP	EA24h	EA48h
		-----%----			%		g	cm	.....%.....	
1	MV	82d	41d	11,20c	78c	11,46e	0,030d	5,68c	79a	53b
2	MV	92c	68c	13,69b	98a	15,95c	0,030d	5,66c	62 b	55b
3	MV	93bc	62c	13,60b	86b	13,54d	0,032d	5,66c	74 a	75a
4	BV	0f	0e	0d	0e	0g	0e	0e	0c	0c
5	BV	8e	0e	0,66d	11d	1,58f	0,09c	4,82d	0c	0c
6	BV	0f	0e	0d	0e	0g	0e	0e	0c	0c
7	AV	97ab	91a	15,81a	97a	24,20a	0,28b	11,42b	76a	85a
8	AV	99a	83b	15,63a	94a	22,25b	0,26b	12,75a	59b	58b
9	AV	96abc	86a	15,32a	96a	23,54a	0,30a	12,6a	63b	57b
<b>CV (%)</b>		4.94	11.19	6.29	5.13	4.51	11.58	8.42	11.56	17.61

Letras diferentes na coluna diferem estatisticamente pelo teste *t* a 5% de probabilidade.

AV= Alto vigor; MV= Médio vigor; BV= Baixo vigor; CV= coeficiente de variação.

Os lotes de sementes de gergelim com maiores porcentagens de germinação foram 7, 8 e 9, todos com germinação acima de 96%, enquanto os lotes 1, 2 e 3 tiveram germinação entre 82 a 93%, e os lotes 4, 5 e 6 apresentaram as menores porcentagens, variando de ausência a 8% de germinação (TABELA 1).

Os valores do teste de germinação indicam a viabilidade das sementes, mas não mostram seu potencial fisiológico, pois é realizado em condições controladas (MARCOS FILHO, 2015; LARIOS *et al.*, 2020a). Esta é a razão para os testes de vigor serem utilizados em conjunto com o teste de germinação, pois avaliam a capacidade das sementes de gerar plântulas normais em condições adversas e são mais sensíveis na detecção de diferenças entre lotes com mesma porcentagem de germinação (SOUSA *et al.*, 2018).

Ainda na Tabela 1, encontram-se os resultados dos testes de vigor utilizados no presente estudo. Percebe-se que os testes estão fortemente correlacionados, no entanto, cada teste tem uma sensibilidade diferente na identificação de alterações entre as amostras.

Análise exploratória com Análise de Componentes Principais (ACP) dos dados fisiológicos foi realizada para confirmar a similaridade dos testes de vigor (APÊNDICE A). De acordo com a ACP, as variáveis que melhor descreveram os dados foram germinação, primeira contagem de germinação e a emergência de plântulas.

As sementes foram classificadas em três categorias de vigor com base nos resultados dos testes fisiológicos da ACP: alto, médio e baixo vigor. Essa classificação auxilia na tomada de decisão de seleção de sementes. Sementes de média e baixa qualidade resultarão em plantas menos desenvolvidas e produtivas. Como resultado, neste estudo, os lotes de sementes 7, 8 e 9 foram designados como sementes de alto vigor (AV), enquanto os lotes 1, 2 e 3 foram designados como sementes de médio vigor (MV) e os lotes 4, 5 e 6 foram designados como sementes de baixo vigor (BV) (TABELA 1).

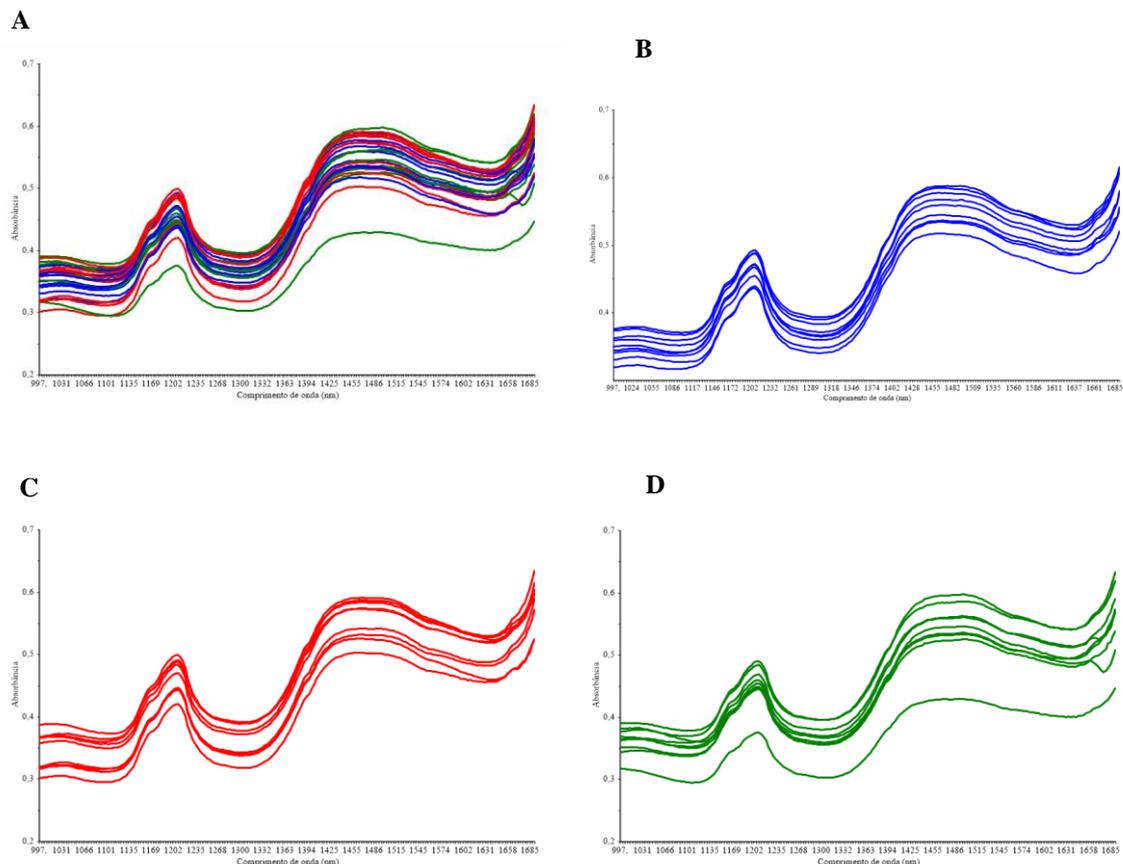
## **6.2 Imagens dos espectros de absorção no infravermelho próximo em sementes de gergelim**

As médias dos espectros obtidos a partir da análise espectral das sementes de gergelim são mostradas na Figura 1A. As Figuras 1B, C e D mostram os espectros separados pela classificação do vigor em alto, médio e baixo vigor, respectivamente, com base nos resultados da Tabela 1. As regiões espectrais de 901 a 992 nm e 1695 a 1700 nm foram excluídas do estudo por serem ruidosas e não apresentarem informações espectrais.

Em geral, verificaram-se comportamentos semelhantes entre as sementes, o que permite inferir que a composição química das amostras de sementes de gergelim com diferentes

níveis de vigor é similar, haja vista que há pequenas diferenças entre seus espectros, no entanto, estas diferenças puderam ser demonstradas com o auxílio da quimiometria. Os espectros mostraram diferenças de linha de base (altura espectral) causadas por espalhamento de luz. Isso é um comportamento comum em espectros brutos (XAVIER *et al.*, 2019). Todavia, as diferenças de linha de base podem ser facilmente removidas pelo pré-tratamento espectral (AGELET *et al.*, 2012) e as diferenças entre as médias dos espectros para alto, médio e baixo vigor podem ser detectadas por modelos de classificação baseados em análise multivariada e quimiometria.

Figura 1 - Espectros brutos de reflectância difusa das sementes de gergelim sem pré-tratamento, obtidos na faixa espectral de 900 a 1700 nm em espectrofotômetro portátil (NIR), para todos os níveis de vigor (A), alto vigor (B), médio vigor (C) e baixo vigor (D).



### 6.3 Modelos quimiométricos de classificação dos lotes de sementes de gergelim

Os resultados dos modelos de regressão PLS-DA com as amostras de treinamento e os diferentes pré-tratamentos revelaram que o pré-tratamento com segunda derivada, segunda derivada + MSC e *Smoothing* + primeira derivada tiveram os melhores desempenhos para as

amostras de sementes com alto, médio e baixo vigor, respectivamente, com os maiores valores de  $R^2$  (0,99) em todos os níveis de vigor para a calibração dos modelos, em comparação com os outros pré-tratamentos e os espectros sem tratamento (APÊNDICE A). A validação interna teve bom ajuste, com  $R^2$  de 0,73, 0,65, e 0,96 para as amostras de sementes com alto, médio e baixo vigor, respectivamente. Como esperado, os valores de  $R^2$  da validação são menores que o modelo de calibração, devido as amostras serem naturais e pelo estudo está correlacionando a variabilidade dos dados espectrais com apenas dois valores numéricos.

Para avaliar o potencial de predição do modelo PLS-DA, amostras que não foram usadas para a validação cruzada, foram usadas para validação externa. Os resultados dos parâmetros de avaliação do modelo em relação às classificações das amostras de lotes de alto, médio e baixo vigor para a validação cruzada (treinamento) e validação externa (teste) estão apresentados na Tabela 2.

Os lotes de alto e baixo vigor não tiveram amostras classificadas como falso positivo na validação cruzada, exceto para o lote de médio vigor, com um total de 17%; já, para a validação externa, não ocorreram amostras classificadas como falso positivo (TABELA 2). Em relação à taxa de falso negativo, não ocorreram amostras classificadas como falso negativo em ambas as validações. O maior número de erros de classificação neste modelo para a taxa de falso positivo pode ser devido à maior proximidade entre as amostras de sementes de alto e médio vigor, conforme observado na Tabela 1.

Todas as amostras foram classificadas com 100% de sensibilidade, tanto para a validação cruzada quanto para a validação externa. Para os testes de especificidade e eficiência, os valores foram iguais, no qual os lotes de alto, médio e baixo vigor tiveram 100, 83 e 100% de acertos para a validação cruzada, respectivamente; para a validação externa, todas as amostras foram classificadas com 100% de especificidade e eficiência.

Tabela 2 - Parâmetros de classificação obtidos do modelo PLS-DA para classificação das sementes de gergelim.

Parâmetros	Validação cruzada (n=19)			Validação externa (n=8)		
	A	M	B	A	M	B
Variáveis latentes	7	6	7	7	6	7
Taxa de falso positivo	0	17	0	0	0	0
Taxa de falso negativo	0	0	0	0	0	0
Sensibilidade	100	100	100	100	100	100
Especificidade	100	83	100	100	100	100
Eficiência	100	83	100	100	100	100

A, M e B = Alto, médio e baixo vigor respectivamente; n= número de amostras.

Esses resultados indicam que o modelo proposto teve bom desempenho, visto que a validação cruzada é um dos testes que comprova o ajuste da calibração do modelo

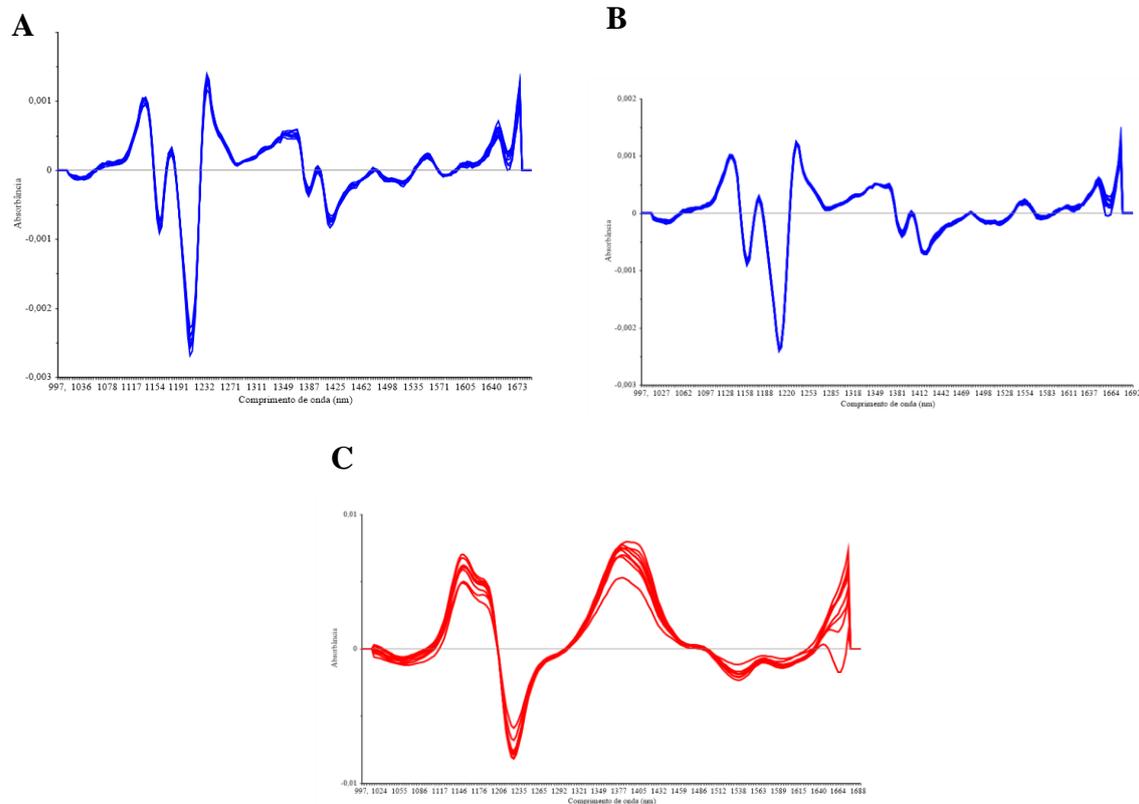
(VASCONCELOS *et al.*, 2018). Portanto, a porcentagem de acertos demonstra que os lotes foram classificados de acordo com os níveis de vigor. Simeone, Ribeiro e Trindade (2019) obtiveram resultados semelhantes ao presente estudo, usando o modelo PLS-DA para classificar sementes de milho haploides com 100% de eficiência.

Xavier *et al.* (2019) obtiveram resultados semelhantes ao do presente estudo usando o modelo PLS-DA para diferenciar espécies de sementes de amaranto com uma precisão de mais de 92%. Vasconcelos *et al.* (2018) usaram modelos de validação cruzada para classificar sementes de girassol com e sem pericarpo, por meio do método PLS-DA, e observaram que as sementes diferiram no modelo e tiveram um bom desempenho, com média de 99 e 100% de precisão na validação cruzada em sementes com e sem pericarpo e 100% de acerto na validação externa para ambos os tratamentos. Mayrinck *et al.* (2020) utilizaram modelos de validação cruzada para classificar sementes de algodão maceradas, construídas a partir do método PLS - DA, e tiveram 95% de acertos para sementes de alto vigor, 100% de acertos para as de baixo vigor (com línter), 100% de acertos para sementes de alto vigor e 91,7% de previsões corretas para as sementes de baixo vigor (sem línter).

#### **6.4 Interpretação dos espectros de absorção no infravermelho próximo em sementes de gergelim**

Os espectros que receberam tratamentos dos dados (FIGURA 2), apresentaram faixas de absorbância mais pronunciadas e mais densas, em comparação com os da Figura 1. Mesmo após o pré-tratamento, pequenas diferenças visuais foram observadas entre os espectros das amostras de alto e médio vigor das sementes de gergelim, enquanto os espectros das amostras de baixo vigor foram mais distintos em relação às demais amostras. Portanto, é importante entender a interpretação das bandas de absorção e os seus respectivos grupos funcionais.

Figura 2 - Espectros de infravermelho próximo de sementes de gergelim com diferentes níveis de vigor, com os pré-tratamentos segunda derivada Savitzky-Golay, segunda derivada Savitzky-Golay + MSC e *Smoothing* + primeira derivada Savitzky-Golay, obtidos na faixa espectral de 900 a 1700 nm em espectrofotômetro portátil (NIR), para as sementes de alto vigor (A), médio vigor (B) e baixo vigor (C), respectivamente.



A Tabela 3 mostra os grupos funcionais encontrados nas bandas espectrais das sementes de gergelim com diferentes níveis de vigor e interpretados de acordo com o Guia Prático e Atlas Espectral para Espectroscopia Interpretativa do Infravermelho Próximo (WORKMAN JÚNIOR e WEYER, 2012). Observa-se que os comprimentos de onda com maior predominância são verificados nas regiões correspondentes aos hidrocarbonetos alifáticos (1040, 1209 e 1397 nm), os quais são obtidos pela degradação química de óleo ou gordura. Esse resultado pode ser explicado pelo fato de que ácidos graxos representam cerca da metade do peso da semente de gergelim e, portanto, é mais facilmente detectável (ANTONIASSI *et al.*, 2013; FREITAS *et al.*, 2018; VALERIANO *et al.*, 2019). De acordo com Hacisalihoglu *et al.* (2019), a presença do metileno (grupo funcional) encontrado nos comprimentos de onda de 1040 e 1209 nm se deve ao fato de ser o componente molecular fundamental dos ácidos graxos saturados e insaturados, responsáveis pela maior parte do conteúdo dos óleos vegetais.

Tabela 3 - Bandas de absorção presentes nas sementes de gergelim obtidas pelo espectrofotômetro NIR.

Comprimento de onda (nm)			Grupo funcional	Tipo de composto
AV	MV	BV		
1040	1040	-	C-H Metileno C-H, associado com R alifático linear R(CH <sub>2</sub> ) NR	Hidrocarboneto alifático
1170	1170	1170	C-H Aceno (.HC=CH)	Alceno, polieno
1206	-	1206	O-H da água perto de 0°C	Água
1209	1209	1209	C-H Metileno (. CH <sub>2</sub> )	Hidrocarboneto alifático
1397	1397	1397	C-H Metil C-H, associado com ramificação RC alifático (CH <sub>3</sub> ) <sub>3</sub> ou RCH(CH <sub>3</sub> ) <sub>2</sub>	Hidrocarboneto alifático
1446	1446	-	C-H Aromático (ArC-H)	Hidrocarboneto aromático
1449	1449	1449	C=O (Carbonyl >C=O)	Cetonas e aldeídos

AV= Alto vigor; MV= Médio vigor; BV= Baixo vigor.

No presente estudo, o grupo funcional alceno representado pela banda de absorção de 1170 nm indica presença de lignina e estiramento do segundo sobretom de C-H (BURNS e CIURCZAK, 2008; WORKMAN JÚNIOR e WEYER, 2012). Em relação aos compostos constituintes, as fibras que constituem a parede celular das sementes de gergelim variam de 2,94 a 3,49% (ANTONIASSI *et al.*, 2013). Outros estudos já demonstraram o uso da espectroscopia no infravermelho próximo (NIR) em conjunto com a quimiometria para a predição de fibras. Por exemplo, Meenu *et al.* (2021) usaram o NIR para prever o conteúdo de β- glucana (fibras) em aveia, obtiveram resultados satisfatórios, com R<sup>2</sup> de 0,789 e 0,677 para sementes de aveia descascadas e nuas, respectivamente.

O teor de água das sementes de gergelim variou de 5,2 a 6,2%, o que justifica o aparecimento da banda em 1206 nm, que representa a absorção de água. A banda de absorção em 1397 nm corresponde ao grupo funcional C-H Metil, que está relacionado com a ramificação alifática RC. As regiões de comprimento de onda de 1446 e 1449 nm correspondem a hidrocarbonetos aromáticos, que estão relacionados à ligação C-H e cetonas, que estão relacionadas à ligação C=O, respectivamente (TABELA 3). Devido à sobreposição de bandas espectrais, algumas bandas não possuem a identificação dos componentes químicos específicos. Outros autores já mencionaram as dificuldades em entender as bandas espectrais sobrepostas - William e Norris (2001) notaram dificuldades na interpretação precisa dos valores de absorbância para misturas complexas, como sementes, assim como Kusumaningrum *et al.* (2017) também relataram dificuldades em reconhecer bandas sobrepostas em espectros de sementes de soja. Esses problemas podem ser causados pela alta concentração de componentes

químicos nas sementes. No presente estudo, os resultados podem estar ligados à composição química das sementes de gergelim.

## **7 CONCLUSÃO**

A combinação da espectroscopia no infravermelho próximo com a análise multivariada foi eficiente para avaliar o vigor de lotes de sementes de gergelim.

Os modelos de regressão PLS-DA com os pré-tratamentos segunda derivada, segunda derivada + MSC e *Smoothing* + primeira derivada tiveram melhor acurácia de classificação para os níveis de alto, médio e baixo vigor dos lotes, respectivamente.

O MicroNIR portátil mostrou-se eficiente para desenvolver modelos de classificação multivariada que permitiu a identificação entre lotes de sementes de gergelim com diferentes níveis de vigor.

Os testes de validação cruzada e externa revelaram que lotes de sementes de gergelim com diferentes níveis de vigor foram classificadas adequadamente com mais de 80% de eficiência. Esses resultados indicam que a técnica é sensível para avaliar o vigor de sementes de gergelim, além de rápida e não destrutiva.

## 8 CAPITULO II: ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO E ANÁLISE MULTIVARIADA NA AVALIAÇÃO DO VIGOR DE SEMENTES DE SORGO

### RESUMO

A semente é o primeiro insumo agrícola na produção vegetal. Portanto, investir em técnicas rápidas e seguras relacionadas à qualidade de sementes é a primeira e mais importante fronteira para aumentar a produção agrícola. Neste estudo, avaliou-se o uso da espectroscopia no infravermelho próximo (NIR) combinado à análise multivariada para classificar o vigor de lotes de sementes de sorgo. Foram utilizados quatro lotes de semente de sorgo, com potenciais fisiológicos distintos, sendo avaliados pelos testes de germinação, primeira contagem de germinação, índice de velocidade de germinação, emergência de plântulas, índice de velocidade de emergência, massa seca de plântulas, comprimento de plântulas, condutividade elétrica e envelhecimento acelerado. Espectros das sementes obtidos no NIR foram analisados por quimiometria e análise multivariada, utilizando análise discriminante por regressão dos mínimos quadrados parciais (PLS-DA). Os principais espectros de NIR das sementes de sorgo ocorrem nas regiões entre 997 a 1452 nm. A interpretação dos espectros de NIR mostraram que a água foi o composto mais importante nas sementes de sorgo. O uso da espectroscopia no infravermelho próximo combinado à quimiometria é eficiente para classificação dos lotes de sementes de sorgo com diferentes níveis de vigor. O modelo de regressão PLS-DA com validação cruzada usando o pré-tratamento segunda derivada + SNV teve melhor classificação dos níveis de alto e baixo vigor das sementes de sorgo. Os testes de validação externa indicaram que sementes de sorgo com alto e baixo vigor foram classificadas com 100% de acertos. Os resultados indicam que o NIR combinado à análise multivariada pode ser utilizado para a determinação eficaz do vigor de sementes de sorgo.

**Palavras-chave:** *Sorghum bicolor*; NIR; quimiometria; PLS-DA.

## CHAPTER II: NEAR INFRARED SPECTROSCOPY AND MULTIVARIATE ANALYSIS IN THE ASSESSMENT OF SORGHUM SEED VIGOR

### ABSTRACT

The seed is the first agricultural input in plant production. Therefore, investing in fast and safe techniques related to seed quality is the first and most important frontier to increase agricultural production. In this study, we evaluated the use of near-infrared (NIR) spectroscopy combined with multivariate analysis to classify the vigor of sorghum seed lots. Four sorghum seed lots were used, with different physiological potentials, being evaluated by germination tests, first germination count, germination speed index, seedling emergence, emergence speed index, seedling dry mass, seedling length, electrical conductivity and accelerated aging. Seed spectra obtained from the NIR were analyzed by chemometrics and multivariate analysis, using discriminant analysis by partial least squares regression (PLS-DA). The main NIR spectra of sorghum seeds occur in the regions between 997 and 1452 nm. Interpretation of NIR spectra showed that water was the most important compound in sorghum seeds. The use of near infrared spectroscopy combined with chemometrics is efficient for classifying sorghum seed lots with different vigor levels. The cross-validated PLS-DA regression model using the second derivative + SNV pretreatment had better classification of high and low vigor levels of sorghum seeds. The external validation tests indicated that sorghum seeds with high and low vigor were classified with 100% correct answers. The results indicate that NIR combined with multivariate analysis can be used for the effective determination of sorghum seed vigor.

**Keywords:** *Sorghum bicolor*; NIR; chemometrics; PLS-DA.

## 9 INTRODUÇÃO

O sorgo (*Sorghum bicolor* L. Moench.) é um dos principais cereais produzidos no mundo e é utilizado na composição de alimentos (importante na segurança alimentar), ração animal e combustível (LIN *et al.*, 2021). Ainda assim, muitos dos insumos empregados na cultura são utilizados de forma amadora, como o uso de sementes com baixa viabilidade e vigor para o plantio. Em razão disso, é comum encontrar diferenças no potencial fisiológico das sementes entre espécies, cultivares e até mesmo entre lotes de sementes (SARMENTO *et al.*, 2020).

A semente é o primeiro e mais importante insumo agrícola, pois determina em grande parte o uso de outros insumos importantes (água, adubação, manejo fitossanitário) que podem melhorar a produtividade. Nesse contexto, o vigor da semente é um dos componentes-chave da qualidade da semente e refere-se ao potencial de uma semente germinar rápida e uniformemente sob várias condições de campo (WANG *et al.*, 2021). Sementes de alto vigor têm vantagens de crescimento e potencial de rendimento, podem melhorar a taxa de emergência no campo e têm forte resistência ao estresse abiótico (WANG *et al.*, 2020).

Existem vários métodos para determinar o vigor das sementes, como envelhecimento acelerado (LIMA *et al.*, 2019), condutividade elétrica (SOUSA *et al.*, 2018), testes fisiológicos e análise de imagens (MEDEIROS *et al.*, 2022). Estes testes, embora precisos e eficazes na detecção do vigor, apresentam limitações em seu uso na produção em larga escala, por serem testes demorados e trabalhosos, que demandam muita mão de obra e causam danos as sementes (testes destrutivos).

Portanto, há necessidade de desenvolver e implementar testes rápidos, eficientes e não destrutivos para avaliar o vigor das sementes, especialmente nos moldes da agricultura que vivemos atualmente – agricultura 4.0, que preconiza a rapidez e a eficiência na obtenção de informações para a melhor tomada de decisão agrônômica e ambiental. Isso ajudará a tornar a agricultura mais sustentável (redução do tempo de avaliação e o custo de mão de obra e sementes), contribuindo para diversos setores dentro da cadeia produtiva da cultura, incluindo as agroindústrias que utilizam o sorgo para alimentos, rações e biocombustíveis, bem como produtores e empresas de sementes que trabalham para melhorar a qualidade do produto (LIN *et al.*, 2021).

A espectroscopia no infravermelho próximo (NIR) tem sido amplamente utilizada combinada à métodos de análise multivariada (quimiometria) em áreas diversas, como alimentos (XU; ZHOU; CHEN, 2017; MBURU *et al.*, 2021), agricultura (RIBEIRO *et al.*, 2021; MEENU *et al.*, 2022) e farmacêuticas (CHEN *et al.*, 2019; MAYR *et al.*, 2021), por ser rápida,

fácil de operar e não destrutiva. Na tecnologia de sementes, o NIR tem sido usado para avaliar o vigor de sementes (WANG *et al.*, 2020; MAYRINCK *et al.*, 2020; LARIOS *et al.*, 2020), viabilidade (KUSUMANINGRUM *et al.*, 2017; MUKASA *et al.*, 2019), identificar cultivares e variedades de sementes (VASCONCELOS *et al.*, 2018; CHEN *et al.*, 2019; YASMIN *et al.*, 2019) e quantificar teor de umidade e nutrientes (SHI; YU, 2017).

O NIR identifica a composição química de materiais biológicos e os grupos funcionais são identificados, como C-H, O-H e N-H com base na absorção de radiação infravermelha por ligações entre átomos de luz que resultam em sobretons e bandas de combinação detectáveis na região de comprimento de onda de 780 – 2500 nm (MUKASA *et al.*, 2019; MEDEIROS *et al.*, 2022). No entanto, para algumas culturas, como o sorgo, tais informações quimiométricas precisam ser melhor elucidadas (LIN *et al.*, 2021).

Com base nesse pressuposto, investigou-se o uso do NIR combinado à análise multivariada para classificar o vigor de lotes de sementes de sorgo.

## **10 MATERIAL E MÉTODOS**

### **10.1 Local do estudo e material utilizado**

A caracterização fisiológica das sementes de sorgo foi realizada no Laboratório de Análise de Sementes, pertencente ao Departamento de Fitotecnia do Centro de Ciências Agrárias da Universidade Federal do Ceará (UFC). A análise de espectroscopia no infravermelho próximo das sementes foi realizada com auxílio do NIR (DLP® NIRscan™ Nano Evaluation Module, Texas Instruments) no Laboratório de Química Analítica da Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB). Quatro lotes de sementes de sorgo foram utilizados para as análises fisiológicas e espectroscópicas.

### **10.2 Condução e delineamento experimental**

O estudo foi realizado em duas partes, na primeira, foram realizados os testes fisiológicos para caracterização do perfil dos lotes de sementes de sorgo, e na segunda, foi realizada a análise de espectroscopia no infravermelho próximo das amostras das sementes com o NIR. O delineamento experimental utilizado para classificar os lotes foi o inteiramente casualizado, com quatro repetições de 50 sementes.

Inicialmente, os perfis dos lotes foram caracterizados e diferenciados por diversos uma série de testes, incluindo teor de água, testes de germinação e de vigor (primeira contagem de germinação, índice de velocidade de germinação, emergência de plântulas, índice de

velocidade de emergência, comprimento de plântulas, massa seca de plântulas, condutividade elétrica e envelhecimento acelerado).

### **10.3 Caracterização fisiológica dos lotes de sementes de sorgo**

#### ***10.3.1 Determinação do teor de água (TA)***

Foram realizados pelo método da estufa a  $105 \pm 3$  °C por 24 h, utilizando-se duas subamostras de 5g de sementes por lote. Os resultados foram expressos em porcentagem com base no peso úmido (BRASIL, 2009).

#### ***10.3.2 Germinação (G) e primeira contagem de germinação (PC)***

As sementes foram submetidas à assepsia com hipoclorito de sódio a 2% por três minutos, lavadas em água destilada e, em seguida, secas com papel toalha e armazenadas em ambiente controlado até o início do estudo. As sementes foram semeadas em rolo de papel umedecido com água destilada, na proporção de 2,5 vezes o peso do papel seco, e mantidas em câmaras de germinação do tipo BOD (*Biochemical Oxygen Demand*) a 25 °C com fotoperíodo de 12 h de luz e 12h de escuro, com contagem inicial de germinação no quarto e a contagem final no décimo dia.

Os resultados foram expressos em porcentagem média de plântulas normais por lote (BRASIL, 2009).

#### ***10.3.3 Índice de velocidade de germinação (IVG)***

Foi realizado simultaneamente com o teste de germinação, com contagens diárias de plântulas normais, até o décimo dia após a instalação do teste. O IVG foi calculado pelo somatório das razões do número de plântulas germinadas diariamente, em comparação com o número de dias em que ocorreu a germinação (MAGUIRE, 1962).

#### ***10.3.4 Emergência (E) e Índice de velocidade de emergência (IVE)***

As sementes foram semeadas em canteiro de areia de 1,0 m de largura  $\times$  10 m de comprimento, com profundidade de semeadura de 2,0 cm. A umidade do substrato foi controlada com irrigação manual. As avaliações foram feitas diariamente, computando-se o número de plântulas que apresentaram a parte aérea emergida até o décimo dia após a semeadura. O resultado foi calculado em porcentagem de acordo com a fórmula proposta por Maguire (1962).

### **10.3.5 Comprimento de plântulas (CP)**

Ao término do teste de germinação, o comprimento de plântulas foi avaliado por meio de análise de imagens. Em cada lote, foram selecionadas 10 plântulas normais. As plântulas de sorgo foram transferidas para papel A3 de coloração branca para a captura de imagens digitais, com auxílio de *scanner*, modelo HP *Scanjet* 2004, montado de maneira invertida dentro de caixa de alumínio. As imagens capturadas foram salvas com resolução digital de 300 dpi, em formato JPEG, e analisadas manualmente usando *software* ImageJ® (*Image Processing and Analysis in Java*) (FERREIRA e RASBAND, 2012). Os resultados obtidos foram expressos em cm plântula<sup>-1</sup>.

### **10.3.6 Massa seca de plântulas (MS)**

Após a medição do comprimento das plântulas, elas foram acondicionadas em sacos de papel, levadas à estufa de circulação de ar forçada à  $65 \pm 3$  °C por 72 horas. A massa seca foi então determinada em balança analítica de precisão (0,0001 g), e os resultados foram representados em g plântula<sup>-1</sup> (KRZYZANOWSKI *et al.*, 2020).

### **10.3.7 Condutividade elétrica (CE)**

As sementes foram colocadas em copos de plástico descartável com capacidade para 200 mL, contendo 75 mL de água destilada e foram mantidos à 25 °C por 24 h em BOD (SOUSA *et al.*, 2018). Em seguida, foram realizadas as leituras com o auxílio do medidor de condutividade Marconi®, modelo MA-521. Os resultados foram expressos em  $\mu\text{S cm}^{-1} \text{ g}^{-1}$  (KRZYZANOWSKI *et al.*, 2020).

### **10.3.8 Envelhecimento acelerado (EA)**

O envelhecimento acelerado das sementes de sorgo foi realizado pelo método tradicional, com umidade relativa em torno de 100%. As sementes foram dispostas em tela de alumínio sobrepostas em caixas plásticas (11 × 11 × 3,5 cm) tipo *gerbox*®, contendo 40 mL de água destilada no fundo e acondicionadas em câmara de germinação tipo BOD a 41 °C por 24 e 48 horas.

#### 10.4 Análise estatística para a caracterização do perfil dos lotes de sementes de sorgo

Os resultados obtidos foram submetidos à análise de variância pelo teste F ( $p \leq 0,05$ ). Em caso de significância, os valores médios dos fatores quantitativos foram comparados pelo teste  $t$  ( $p \leq 0,05$ ), utilizando o *software* estatístico SISVAR<sup>®</sup> (FERREIRA, 2011).

#### 10.5 Obtenção dos espectros por espectroscopia NIR em sementes de sorgo

Os espectros na região do infravermelho próximo foram obtidos em amostras de 4,5 g de sementes de sorgo inteiras homogeneizadas para cada lote. Utilizou-se o espectro NIR, na faixa de 900 nm a 1700 nm, com um incremento/largura de 7 nm, 50 varreduras e resolução digital de 228 comprimentos de onda, em ambiente com temperatura de 20 °C. A leitura dos espectros, em triplicatas, foi realizada em Becker de 5 mL contendo as amostras, sendo utilizado como espectro de referência (branco) o Teflon. As médias das leituras dos espectros totalizaram 12 amostras espectrais que foram utilizadas para realização dos pré-tratamentos dos dados, análise do comportamento espectral e construção do modelo de classificação, utilizando o pacote do *software* *The Unscrambler*<sup>®</sup> X versão 10.4 (Camo Analytics, Oslo, Norway).

#### 10.6 Análise espectral em sementes de sorgo

Foram utilizados pré-tratamentos nos dados espectrais para corrigir efeitos de espalhamento de luz e as demais influências físicas que não estejam relacionadas com a composição química das sementes, como correção multiplicativa de espalhamento (*Multiplicative Scatter Correction* – MSC), variação de padrão normal (*Standard Normal Variate* – SNV), primeira e segunda derivada Savtizy-Golay com polinômio de segunda ordem e janela de 11 pontos. Após a aplicação dos pré-tratamentos, as amostras foram classificadas de acordo com o nível de vigor para cada pré-tratamento dos espectros utilizando a técnica de classificação supervisionada e o algoritmo Análise Discriminante por Mínimos Quadrados Parciais (*Partial Least Squares - Discriminant Analysis*, PLS-DA).

As amostras das sementes de sorgo foram divididas em dois conjuntos para a construção dos modelos de classificação. No primeiro conjunto, foram utilizados 70% das amostras de sementes para treinamento (Calibração) e, no segundo conjunto, 30% das amostras foram utilizadas para testar o modelo (Validação). Desta maneira, foi otimizado um modelo por validação cruzada, sendo as classes do vetor Y as variáveis dependentes e os espectros obtidos foram as variáveis independentes do vetor X (ABDI, 2003).

Nos modelos PLS-DA, os valores de classes -1 e 1 foram atribuídas as amostras para calibração das sementes pertencentes aos níveis de baixo e alto vigor, respectivamente.

O conjunto de treinamento foi utilizado para obtenção e escolha do número de variáveis latentes - PLS (VLs), um parâmetro importante do modelo PLS-DA. A escolha do melhor pré-tratamento foi realizada com base no menor erro do modelo e o maior número de acertos de validação.

Para avaliar o desempenho e eficiência do modelo de classificação desenvolvido foi criada a Matriz de Confusão, descrita por Botelho et al. (2015), para avaliar os seguintes parâmetros: taxas de falso positivo (TFP), falso negativo (TFN), taxa de sensibilidade (SEN), taxa de especificidade (ESPEC), a taxa de eficiência (TEF) em %.

A taxa de falso positivo é a porcentagem de amostras negativas que foram classificadas como amostras positivas, e é calculada como a relação entre o número absoluto de falsos positivos (FP) e a soma do número absoluto de erros falsos positivos (FP) e verdadeiros negativos (VN) multiplicada por 100, representada pela equação 1:

$$\mathbf{TFP = FP/(FP+VN) \times 100} \quad \mathbf{(1)}$$

Da mesma forma, a taxa de falso negativo é a porcentagem de amostras positivas que foram classificadas como amostras negativas, sendo calculada como a relação entre o número absoluto de falsos negativos (FN) e a soma do número absoluto de erros falsos negativos (FN) e verdadeiros positivos (VP) multiplicada por 100, representada pela equação 2:

$$\mathbf{TFN = FN/(FN+VP) \times 100} \quad \mathbf{(2)}$$

A sensibilidade é a porcentagem de amostras positivas que foram classificadas corretamente como sendo dessa classe, sendo calculada como a razão entre o número absoluto de verdadeiros positivos (VP) e a soma do número absoluto de verdadeiros positivos (VP) e dos erros falsos negativos (FN) multiplicada por 100, representada pela equação 3:

$$\mathbf{SEN=VP/(VP+FN) \times 100} \quad \mathbf{(3)}$$

A especificidade é a porcentagem de amostras negativas que foram classificadas corretamente pelo modelo como as amostras negativas, sendo calculada pela razão entre o número absoluto de verdadeiros negativos (VN) e a soma do número absoluto de verdadeiros negativos (VN) e dos erros falso positivos (FP) multiplicado por 100, representada pela equação 4:

$$\text{ESPEC} = \text{VN} / (\text{VN} + \text{FP}) \times 100 \quad (4)$$

A eficiência combina todas as informações transportadas pelas taxas de falso positivo e negativo, sendo calculada pela diferença entre o valor de 100% e a soma das taxas de erros falsos negativos (TFN) e falsos positivos (TFP), representada pela equação 5:

$$\text{TEF} = 100 - (\text{TFN} + \text{TFP}) \quad (5)$$

Os pré-tratamentos e os modelos de classificação dos dados foram realizados por software *The Unscrambler*<sup>®</sup> X versão 10.4 (Camo Analytics, Oslo, Norway).

## 11 RESULTADOS E DISCUSSÃO

### 11.1 Caracterização dos lotes de sementes de sorgo

Os valores dos parâmetros da avaliação do vigor das sementes estão na Tabela 4. Os lotes apresentaram diferentes comportamentos quanto ao seu potencial fisiológico. Em todos os lotes, observam-se diferenças ( $p \leq 0,05$ ) para as variáveis germinação, primeira contagem de germinação, índice de velocidade de germinação, índice de velocidade de emergência, condutividade elétrica, massa seca de plântulas e envelhecimento acelerado com 24h. No entanto, para as variáveis emergência de plântulas, envelhecimento acelerado com 48h e comprimento de plântulas, não foi observado diferença estatística.

Tabela 4 - Testes de germinação (G), primeira contagem de germinação (PC), índice de velocidade de germinação (IVG), emergência de plântulas (EP), índice de velocidade de emergência (IVE), condutividade elétrica (CE), massa seca de plântulas (MS), comprimento de plântulas (CP), envelhecimento acelerado de 24 (EA24h) e 48h (EA48h).

Lotes	Vigor	G	PC	IVG	EP	IVE	CE	MS	CP	EA24h	EA48h
		----- % ----			%		$\mu\text{S cm}^{-1}$	g	cm	.....%.....	
1	AV	93bc	93a	11,62bc	91	11,24b	42,69b	0,14c	22,82	98a	98
2	AV	98a	97a	12,27a	97	11,55ab	36,96ab	0,19b	23,32	98a	93
3	BV	89c	83b	11,02c	90	9,24c	31,52a	0,22a	22,94	93b	87
4	AV	96ab	92a	11,90ab	95	11,87a	32,39a	0,20b	22,77	97a	95
CV (%)		3.13	5.18	3.39	3.76	3.53	10.62	5.34	5.88	2.22	5.56

Letras diferentes na coluna diferem estatisticamente pelo teste *t* a 5% de probabilidade.

BV= Baixo vigor; AV= Alto vigor; CV= coeficiente de variação.

Os lotes de sorgo com as maiores porcentagens de germinação foram o 2 e 4, com 98 e 96%, respectivamente, seguidos pelos lotes 1 e 3, que obtiveram 93 e 89%, respectivamente (TABELA 4). O teste de germinação indica a viabilidade das sementes, no entanto, não mostram seu vigor, pois o teste é conduzido em condições controladas (MARCOS FILHO, 2015; LARIOS *et al* 2020). Portanto, se justifica aplicar testes de vigor juntamente com o teste de germinação, por avaliar melhor a capacidade das sementes em gerarem plântulas normais em condições adversas e são mais sensíveis em detectar diferenças entre lotes com mesma porcentagem de germinação (SOUSA *et al.*, 2018).

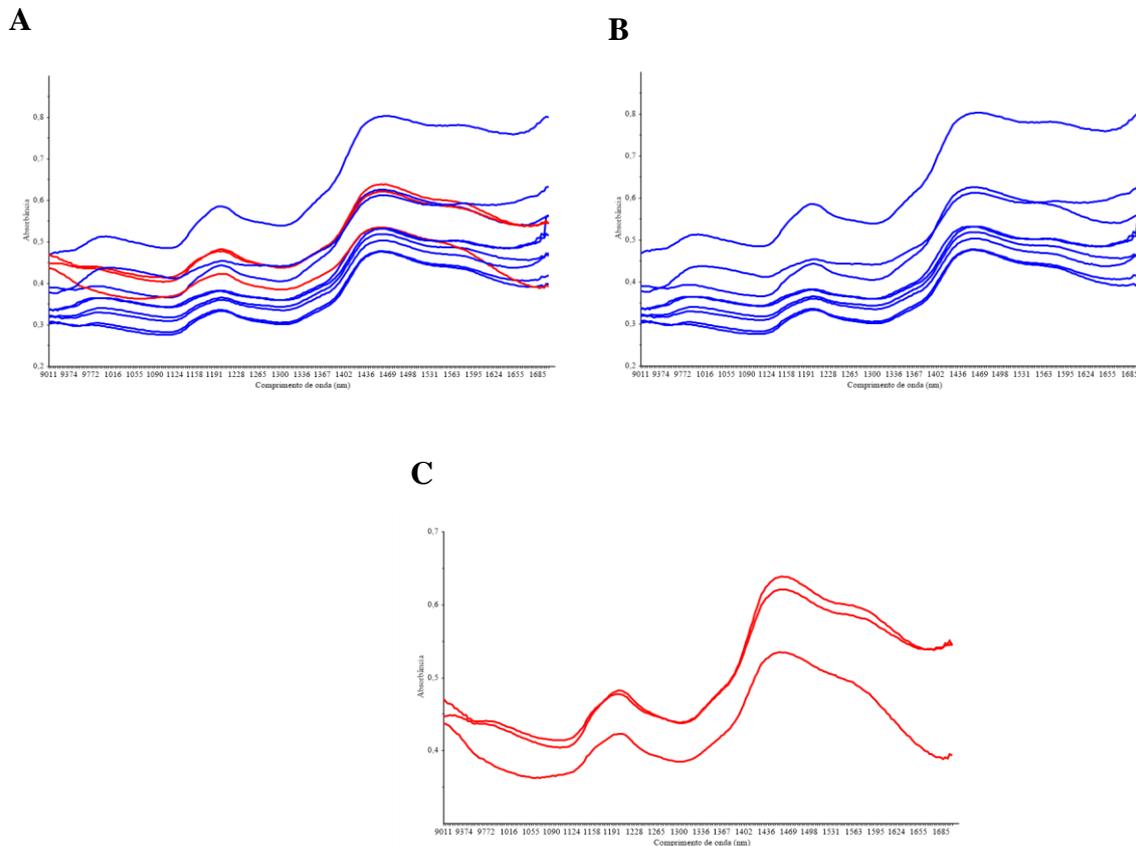
As sementes foram classificadas com base nos resultados dos testes fisiológicos em dois níveis de vigor: alto e baixo vigor. Essa classificação facilita a tomada de decisão para a seleção de sementes. Sementes de baixa qualidade produzirão plantas menos desenvolvidas e produtivas. Portanto, neste estudo, os lotes de sementes 1, 2 e 4 foram classificados como sementes de alto vigor (AV) e o lote 3 como sementes de baixo vigor (BV) (TABELA 4).

## **11.2 Imagens dos espectros de absorção no infravermelho próximo em sementes de sorgo**

As médias dos espectros obtidos a partir da análise espectral das sementes de sorgo são mostrados na Figura 3A. As Figuras 3B e 3C mostra os espectros dos lotes de alto e baixo vigor, respectivamente. A região espectral de 901 a 973 nm e 1583 a 1700 nm foram excluídas do estudo por apresentarem ruídos e não possuem informação espectral. Diferentes intensidades de bandas de absorção foram encontradas nas regiões espectrais de 973-1583 nm. Em geral, os espectros mostraram picos semelhantes entre as amostras de sementes, o que sugere que a composição química das amostras de sementes de sorgo com níveis de vigor é similar, haja vista que não há diferença marcante entre seus espectros.

Os espectros brutos neste estudo apresentaram espalhamento de luz, comportamento comum de dados espectrais brutos de amostras de sementes e pode ser devido à morfologia das sementes e a efeitos indesejáveis causados por fatores externos, como ruído aleatório e mudanças na linha de base (altura espectral) (XAVIER *et al.*, 2019; MEDEIROS *et al.*, 2022). As diferenças de linha de base podem ser facilmente removidas pelo pré-tratamento espectral (AGELET *et al.*, 2012) e as diferenças entre as médias dos espectros para alto e baixo vigor pode ser detectadas por modelos de classificação baseados em análise multivariada e quimiometria, como PCA e PLS-DA.

Figura 3 - Espectros de infravermelho próximo de sementes de sorgo sem pré-tratamento, obtidos na faixa espectral de 900 a 1700 nm em espectrofotômetro portátil (NIR), para todos os níveis de vigor (A) e para as amostras de alto vigor (B) e baixo vigor (C).



### 11.3 Modelos quimiométricos de classificação dos lotes de sementes de sorgo

Os resultados dos modelos das amostras de treinamento, com os diferentes pré-tratamentos, estão organizados no Apêndice B. Os modelos de regressão por PLS-DA mostraram que os pré-tratamentos não melhoraram significativamente a capacidade discriminativa dos modelos, no entanto, para correção das variações de linha de base presentes nos espectros brutos, o modelo com o pré-tratamento com segunda derivada + SNV teve o melhor desempenho, em comparação aos outros pré-tratamentos. Este modelo apresentou os maiores valores de  $R^2$  para a calibração (0,81) e para a validação interna (0,58). Diferentemente do que foi encontrado neste estudo, estudos anteriores mostraram que a aplicação de pré-tratamentos antes do desenvolvimento de modelos pode melhorar a precisão do modelo (QIU *et al.*, 2018).

Resultados da utilização eficiente do NIR em sementes já foram relatados. No entanto, estudos anteriores em sementes avaliaram o uso do NIR em amostras unitárias de sementes, por exemplo, para discriminar a viabilidade de sementes de cipreste de hinoki com

o modelo PLS-DA, que mostrou 97,7–99,2% de acerto de calibração e 94,4–95,4% de validação, usando diferentes métodos de pré-processamento na classificação de sementes viáveis e não viáveis (MUKASA *et al.*, 2019). Em outro estudo, Yasmin *et al.* (2019), ao investigarem o potencial da espectroscopia no infravermelho próximo com transformada de Fourier (FT-NIR) e o modelo PLS-DA em sementes unitárias de melancia triploides para discriminar sementes viáveis e não viáveis de três variedades diferentes, armazenadas por quatro anos em condições controladas, registraram boa precisão dos resultados de classificação para sementes viáveis (87,7%) e sementes não viáveis (82%). Vale ressaltar que a utilização do NIR em cada unidade de semente não é viável para uma triagem de produção em larga escala. Portanto, o presente estudo mostra que a técnica de medição NIR em sementes pode ser abordada tanto em escala laboratorial como industrial.

Para avaliar o potencial de predição do modelo PLS-DA, amostras que não foram usadas para a validação cruzada, foram usadas para validação externa e os resultados dos parâmetros de avaliação do modelo em relação às classificações das amostras de lotes de alto e baixo vigor para a validação cruzada (treinamento) e validação externa (teste) estão apresentados na Tabela 5.

Observa-se na Tabela 5 que não ocorreram amostras classificadas como falso positivo ou falso negativo, dessa forma, a análise dos parâmetros como sensibilidade, especificidade e eficiência resultaram em valores iguais a 100%, tanto para o conjunto de treinamento (validação cruzada) quanto para o conjunto de teste (validação externa), indicando que houve uma excelente classificação dos lotes de sementes de sorgo. Segundo Forina *et al.* (1991), o modelo de melhor ajuste é aquele com valores de sensibilidade e especificidade iguais a 1,0 (100%). Mayrinck *et al.* (2020) afirmaram que os resultados da validação externa indicam a capacidade preditiva do modelo de sementes que não participaram da construção do mesmo modo, indicando que é robusto e capaz de avaliar diferentes amostras.

Tabela 5 - Parâmetros de classificação obtidos do modelo PLS-DA para classificação das sementes de sorgo.

Parâmetros	Validação cruzada (n=8)	Validação externa (n=4)
Variáveis latentes	4	4
Taxa de falso positivo	0	0
Taxa de falso negativo	0	0
Sensibilidade	100	100
Especificidade	100	100
Eficiência	100	100

n= número de amostras.

Estudos anteriores ao utilizarem a análise discriminante de mínimos quadrados parciais (PLS-DA) para classificação de sementes, obtiveram resultados semelhantes ao presente trabalho. Wang *et al.* (2020) ao analisar a viabilidade do NIR para detecção do vigor de sementes de milho doce, registraram precisão de classificação superior a 98% para as amostras de sementes normais, sementes envelhecidas artificialmente e sementes danificadas pelo calor. Kusumaningrum *et al.* (2017) utilizaram a espectroscopia FT-NIR para determinar a viabilidade de sementes de soja com o método PLS-DA e concluíram que o PLS-DA apresenta bom desempenho, com uma precisão próxima a 100% de previsão da viabilidade da soja.

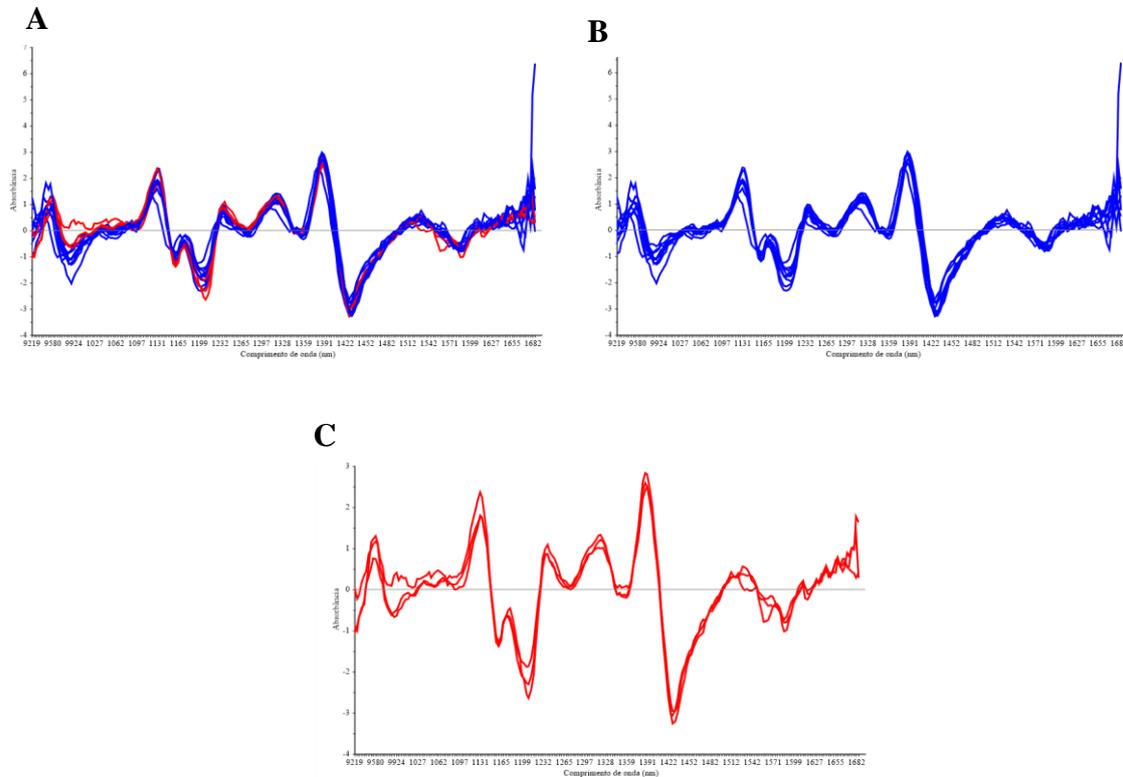
Resultados semelhantes também foram reportados por Mayrinck *et al.* (2020), que utilizando modelos de validação cruzada para classificar sementes de algodão maceradas, construídos a partir do método PLS – DA, registraram para sementes com línter, 95 e 100% de predições corretas para as amostras de alto e baixo vigor, respectivamente, e 100 e 91,7% de predições corretas para sementes sem línter de alto e de baixo vigor.

Embora os dados tenham sido coletados de um número limitado de amostras de sementes de sorgo, em geral, os resultados foram promissores e indicam que o NIR combinado à análise multivariada é um método rápido, seguro e não destrutivo para diferenciar os níveis de vigor em sementes de sorgo.

#### **11.4 Interpretação dos espectros de absorção no infravermelho próximo em sementes de sorgo**

Os espectros que receberam tratamentos (FIGURA 4) apresentaram faixas de absorbância mais pronunciadas e mais densas, em comparação com os da Figura 3. Mesmo com o pré-tratamento pequenas diferenças visuais foram observadas nos espectros das amostras de alto e baixo vigor das sementes de sorgo. Portanto, é importante entender a interpretação das bandas de absorção e os seus respectivos grupos funcionais.

Figura 4 - Espectros de infravermelho próximo de sementes de sorgo com diferentes níveis de vigor, com o pré-tratamentos com segunda derivada Savitzky-Golay + SNV, obtidos na faixa espectral de 900 a 1700nm em espectrofotômetro portátil (NIR), para todos os níveis de vigor (A), alto vigor (B) e baixo vigor (C).



Para as sementes de alto vigor, os comprimentos de onda mais importantes foram registrados em 997, 1199 e 1452 nm. Já, as sementes de baixo vigor, tiveram perda no pico de absorção em relação às de alto vigor, com comprimentos de onda de 1202 e 1452 nm. A identificação das bandas de absorção e os seus respectivos grupos funcionais obtidos no NIR foi realizada segundo Burns e Ciurczak (2008). A região espectral de 997 nm corresponde ao grupo funcional O-H, atribuído a álcoois primários; já o pico em torno de 1199 e 1202 nm corresponde ao grupo funcional O-H, associados à água; o pico em 1452 nm corresponde ao grupo funcional O-H, atribuído a álcoois primários e água. Como resultado, os dados de NIR (na região de 997 a 1452 nm) indicaram que a água foi o principal componente dos grupos funcionais.

Em geral, a presença do grupo funcional O-H (-CH<sub>2</sub>-OH) nos espectros indica a presença de carboidratos nas sementes de sorgo. Segundo Marcos-Filho (2016), os carboidratos são os primeiros compostos armazenados em muitas espécies de sementes, contêm hidrogênio e oxigênio (CH<sub>2</sub>O) na mesma proporção que a água, e sua principal função é fornecer energia ao embrião durante a germinação. Os carboidratos compõem a maior parte das sementes de sorgo, cerca de 70%, com o restante consistindo de lipídios (1,6 a 6%),

proteínas (7 a 15%) (OSEGUERA-TOLEDO *et al.*, 2020) e fibras (4,5 a 26,3%) (BATARIUC; UNGUREANU-IUGA; MIRONEASA, 2021).

No presente estudo, provavelmente a água das sementes (10,3-11,2%) estava adsorvida aos carboidratos. Como a água normalmente tem picos fortes na mesma região do espectro dos carboidratos, isso pode ter contribuído para uma menor expressão das bandas características dos carboidratos (HELL *et al.*, 2015). Dykes *et al.* (2014) também observaram picos espectrais indicando a presença de água nas sementes de sorgo em bandas de absorção em torno de 1450 nm (estiramento OH primeiro sobretom), usando o NIR.

A perda de picos espectrais em sementes de baixo vigor pode estar relacionada à perda de viabilidade nessas sementes. As sementes começam a perder viabilidade quando atingem a maturidade fisiológica, sendo este processo dividido em três fases de declínio. Na primeira fase, a maioria das sementes apresenta alta viabilidade e sem grandes problemas de germinação. Durante a segunda fase, a viabilidade começa a declinar rapidamente e, como resultado, poucas sementes germinam. Durante a terceira fase, poucas sementes permanecem viáveis, na qual o número de sementes não germinadas aumenta (KRZYZANOWSKI *et al.*, 2020).

## **12 CONCLUSÃO**

O uso da espectroscopia no infravermelho próximo combinado à análise multivariada é eficiente para classificar o vigor de lotes de sementes de sorgo.

A classificação dos lotes por vigor pode ser obtida usando o modelo de regressão por PLS-DA com a aplicação da segunda derivada + SNV como pré-tratamento dos espectros.

Os testes de validação externa indicaram que lotes de sementes de sorgo de alto e baixo vigor foram classificadas com 100% de acertos.

## REFERÊNCIAS

- ABDI, H. Partial least square regression (PLS regression). **Encyclopedia for research methods for the social sciences**, [s.l.], v. 6, n. 4, p. 792-795, 2003. DOI: 10.1002/wics.51. Acesso em: 12 jan. 2021.
- AGELET, L. E. *et al.* Feasibility of near infrared spectroscopy for analyzing corn kernel damage and viability of soybean and corn kernels. **Journal of Cereal Science**, [s.l.], v. 55, n. 2, p. 160-165, 2012. DOI: 10.1016/j.jcs.2011.11.002. Acesso em: 26 mar. 2022.
- ALVAREZ, A. M. R. **Avaliação da qualidade fisiológica de sementes de milho e soja por espectroscopia no infravermelho próximo**. Dissertação (mestrado acadêmico) – Universidade Federal de Lavras, Lavras. 2018. 88 p. Disponível em: <http://repositorio.ufla.br/jspui/handle/1/28997>. Acesso em: 02 jan. 2021.
- ALVES, C. Z. *et al.* pH of exudate test in the physiological quality of crambe seeds. **Ciência Rural**, Santa Maria, v. 46, n. 6, p. 1014-1018, 2016. DOI: 10.1590/0103-8478cr20141838. Acesso em: 13 jan. 2021.
- AMBROSE, A. *et al.* Comparative nondestructive measurement of corn seed viability using Fourier transform near-infrared (FT-NIR) and Raman spectroscopy. **Sensors and Actuators B: Chemical**, [s.l.], v. 224, p. 500-506, 2016. DOI: 10.1016/j.snb.2015.10.082. Acesso em: 31 dez. 2020.
- ANTONIASSI, R. *et al.* Influência das condições de cultivo na composição da semente e do óleo de gergelim. **Revista Ceres**, Viçosa, v. 60, n. 3, p. 301-310, 2013. Acesso em: 15 mar. 2022.
- ARRIEL, N. H. C.; BELTRÃO, N. E. de M.; FIRMINO, P. T. **Gergelim: o produtor pergunta, a Embrapa responde**. Brasília: EMBRAPA-CNPQ, 2009. p. 21-38.
- BARBOSA, R. M. *et al.* Image analysis and peanut seeds performance during the production process. **Científica**, Jaboticabal, v. 44, n. 3, p. 412-420, 2016. DOI: 10.15361/1984-5529.2016v44n3p412-420. Acesso em: 19 jan. 2021.
- BARRETO, L. C.; GARCIA, Q. S. Accelerated ageing and subsequent imbibition affect seed viability and the efficiency of antioxidant system in macaw palm seeds. **Acta Physiologiae Plantarum**, [s.l.], v. 39, n. 3, p. 1-8, 2017. DOI: 10.1007/s11738-017-2367-z. Acesso em: 14 jan. 2021.
- BATARIUC, A; UNGUREANU-IUGA, M; MIRONEASA, S. Effects of Dry Heat Treatment and Milling on Sorghum Chemical Composition, Functional and Molecular Characteristics. **Applied Sciences**, [s.l.], v. 11, n. 24, p. 11881, 2021. DOI:10.3390/app112411881. Acesso em: 13 jul. 2022.
- BERRUETA, L. A., ALONSO-SALCES, R. M., HÉBERGER, K. Supervised pattern recognition in food analysis. **Journal of Chromatography A**, [s.l.], v. 1158, n. 1-2, p. 196-214, 2007. DOI: 10.1016/j.chroma.2007.05.024. Acesso em: 14 jan. 2021.

- BOTELHO, B. G. *et al.* Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. **Food Chemistry**, [s.l.], v. 181, p. 31-37, 2015. DOI: 10.1016/j.foodchem.2015.02.077. Acesso em 29 abr.2022.
- BRASIL, Ministério da agricultura, pecuária e abastecimento. **Regras para Análise de Sementes**. Brasília: DNDV/CLAV. 399 p. 2009. Disponível em: [https://www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios/arquivos-publicacoes-insumos/2946\\_regras\\_analise\\_\\_sementes.pdf/view](https://www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios/arquivos-publicacoes-insumos/2946_regras_analise__sementes.pdf/view). Acesso em: 20 mar. 2021.
- BURNS, D. A; CIURCZAK, E. W. **Handbook of near-infrared analysis**. 3. ed. Boca Raton: CRC, 2008, 8008p.
- CARVALHO, S. M. C. *et al.* Viability of *Carica papaya* L. Seeds by the Tetrazolium Test. **Journal of Agricultural Science**, [s.l.], v. 10, n. 2, p. 335-340, 2018. DOI: 10.5539/jas.v10n2p335. Acesso em 13 jan. 2021.
- CASTAN, D. O. C.; GOMES-JUNIOR, F. G.; MARCOS-FILHO, J. Vigor-S, a new system for evaluating the physiological potential of maize seeds. **Scientia Agricola**, [s.l.], v. 75, n. 2, p. 167- 172, 2018. DOI: 10.1590/1678-992X-2016-0401. Acesso em: 19 jan. 2021.
- CHEN, J. *et al.* Rapid and non-destructive analysis for the identification of multi-grain rice seeds with near-infrared spectroscopy. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, [s.l.], v. 219, p. 179-185, 2019. DOI:10.1016/j.saa.2019.03.105. Acesso em: 06 jul. 2022.
- CHEN, X. *et al.* Quality evaluation of decoction pieces of *Rhizoma Atractylodis Macrocephalae* by near infrared spectroscopy coupled with chemometrics. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 221, p. 117169, 2019. DOI: 10.1016/j.saa.2019.117169. Acesso em: 06 jul. 2022.
- CONAB, Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira de grãos**. Brasília, v. 9, n. 11, p. 1-86, 2022. Disponível em: <https://www.conab.gov.br/info-agro/safras/graos>. Acesso em: 18 jan. 2021.
- CRUZ, N. F. F. da S. *et al.* Características e tratos culturais do gergelim (*Sesamum indicum* L.). **Revista Brasileira de Energias Renováveis**, [s.l.], v. 8, n. 4, p. 665- 675, 2019. Disponível em: [https://www.academia.edu/55151367/CARACTER%C3%8DSTICAS\\_E\\_TRATOS\\_CULTURAIS\\_DO\\_GERGELIM\\_Sesamum\\_indicum\\_L\\_](https://www.academia.edu/55151367/CARACTER%C3%8DSTICAS_E_TRATOS_CULTURAIS_DO_GERGELIM_Sesamum_indicum_L_). Acesso em: 18 jan. 2021.
- CRUZ, R. N. *et al.* Adubação orgânica residual no crescimento e produção do gergelim irrigado com água residuária. **Revista Verde de Agroecologia e Desenvolvimento Sustentável**, Mossoró, v. 8, n. 1, p. 257-263, 2013. Disponível em: <https://www.gvaa.com.br/revista/index.php/RVADS/article/view/2012/1592>. Acesso em:18 jan. 2021.
- DYKES, L. *et al.* Prediction of total phenols, condensed tannins, and 3-deoxyanthocyanidins in sorghum grain using near-infrared (NIR) spectroscopy. **Journal of Cereal Science**, [s.l.], v. 60, n. 1, p. 138-142, 2014. DOI: 10.1016/j.jcs.2014.02.002. Acesso em: 15 jul. 2022.

- FAN, Y.; M. A, S.; WU, T. Individual wheat kernels vigor assessment based on NIR spectroscopy coupled with machine learning methodologies. **Infrared Physics and Technology**, [s.l.], v. 105, p. 103213-103213, 2020. DOI: 10.1016/j.infrared.2020.103213. Acesso em: 13 jan. 2021.
- FAOSTAT. FOOD AND AGRICULTURE ORGANIZATION OF UNITED NATIONS. **Crops**. 2019. Disponível em: <<http://www.fao.org/faostat/en/#data/QC>>. Acesso em: 13 jan. 2021.
- FERREIRA, D. F. Sisvar: a computer statistical analysis system. **Ciência e Agrotecnologia**, Lavras, v. 35, n. 6, p. 1039-1042, 2011. Disponível em: <https://doi.org/10.1590/S1413-70542011000600001>. Acesso em: 18 jan. 2021.
- FERREIRA, M. M. C. *et al.* Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, [s.l.], v. 22, n. 5, p. 724-731, set. 1999. Disponível em: <https://doi.org/10.1590/S0100-40421999000500016>. Acesso em: 13 jan. 2021.
- FERREIRA, T.; RASBAND W. **ImageJ user guide**: IJ1.46r. 2012.
- FORINA, M. *et al.* A class modeling technique based on potential functions. **Journal of Chemometrics**, [s.l.], v. 5, p. 435-453, 1991. DOI: 10.1002/cem.1180050504. Acesso em: 27 mar. 2022.
- FREITAS, M. N. de. *et al.* **Composição Química de Três Variedades de Gergelim**. Fortaleza: Embrapa Agroindústria Tropical. 2018. 22p. (Boletim de pesquisa e desenvolvimento / Embrapa Agroindústria Tropical, ISSN 1679-6543; 160). Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1091165/1/BPD18009.pdf>>. Acesso em: 15 mar. 2022.
- GRUNVALD, A. K. *et al.* Predicting the oil contents in sunflower genotype seeds using near-infrared reflectance (NIR) spectroscopy. **Acta Scientiarum**, Maringá, v. 36, n. 2, p. 233-237, 2014. DOI: 10.4025/actasciagron.v36i2.17677. Acesso em: 14 jan. 2021.
- GUIMARÃES, G. C. **Espectroscopia no infravermelho próximo para classificação de sementes de café quanto à qualidade, origem e cultivar**. Tese (doutorado) – Universidade Federal de Lavras, Lavras. 2016.
- HACISALIHOGU, G. *et al.* Protein, weight, and oil prediction by single-seed near-infrared spectroscopy for selection of seed quality and yield traits in pea (*Pisum sativum*). **Journal of The Science of Food and Agriculture**, [s.l.], v. 1, n. 1, p. 1-10, 2020. DOI: 10.1002/jsfa.10389. Acesso em: 26 mar. 2022.
- HART, J. R.; NORRIS, K. H.; GOLUMBIC, C. Determination of the moisture content of seeds by near-infrared spectrophotometry of their methanol extracts. **Cereal Chemistry**, [s.l.], v. 39, p. 94-99, 1962. Disponível em: <https://www.cerealsgrains.org/publications/cc/backissues/1962/Documents/CC1962a11.html>. Acesso em 27 mar. 2021.
- HELL, J. *et al.* A comparison between near-infrared (NIR) and mid-infrared (ATR-FTIR) spectroscopy for the multivariate determination of compositional properties in wheat bran

samples. **Food Control**, [s.l.], v. 60, p. 365-369, 2016. DOI: 10.1016/j.foodcont.2015.08.003. Acesso em 26 mar. 2022.

KARDEŞ, Y. M. *et al.* Biochemical composition of selected lines from sorghum (*Sorghum bicolor* L.) landraces. **Planta**, [s.l.], v. 254, n. 2, p. 1-13, 2021. DOI: 10.1007/s00425-021-03670-9. Acesso em: 27 mar. 2021.

KAUR, B.; SANGHA, M. K.; KAUR, G. Development of Near-Infrared Reflectance Spectroscopy (NIRS) Calibration Model for Estimation of Oil Content in *Brassica juncea* and *Brassica napus*. **Food Analytical Methods**, [s.l.], v. 10, n. 1, p. 227–233, 2017. DOI:10.1007/s12161-016-0572-9. Acesso em 27 mar. 2021.

KONG, W. *et al.* Rice Seed Cultivar Identification Using Near-Infrared Hyperspectral Imaging and Multivariate Data Analysis. **Sensors**, [s.l.], v. 13, n. 7, p. 8916-8927, 2013. DOI: 10.3390/s130708916. Acesso em: 24 mar. 2022.

KRZYŻANOWSKI, F. C. *et al.* **Vigor de sementes: conceitos e testes**. Londrina: ABRATES, 2020. p. 601.

KUMAR, S. *et al.* Infrared spectroscopy combined with imaging: a new developing analytical tool in health and plant science. **Applied Spectroscopy Reviews**, [s.l.], v. 51, n. 6, p. 466-483, 2016. DOI: 10.1080/05704928.2016.1157808. Acesso em: 14 jan. 2021.

KUSUMANINGRUM, D. *et al.* Non-destructive technique for determining the viability of soybean (*Glycine max*) seeds using FT-NIR spectroscopy. **Journal of the Science of Food and Agriculture**, [s.l.], v. 98, n. 5, p. 1734-1742, 2017. DOI: 10.1002/jsfa.8646. Acesso em: 26 mar. 2022.

LARIOS, G. *et al.* Soybean seed vigor discrimination by using infrared spectroscopy and machine learning algorithms. **Analytical Methods**, [s.l.], v. 12, n. 35, p. 4303-4309, 2020. DOI: 10.1039/d0ay01238f. Acesso em: 06 dez. 2021.

LARIOS, G. S. *et al.* Laser-Induced Breakdown Spectroscopy as a Powerful Tool for Distinguishing High- and Low-Vigor Soybean Seed Lots. **Food Analytical Methods**, [s.l.], v. 13, n. 9, p. 1691-1698, 2020. DOI: 10.1007/s12161-020-01790-8. Acesso em: 01 maio 2022.

LEÃO-ARAÚJO, E. F. *et al.* Controlled deterioration test and use of the Seed Vigor Imaging System (SVIS<sup>®</sup>) to evaluate the physiological potential of *crambe* seeds. **Journal of Seed Science**, [s.l.], v. 39, n. 4, p. 393-400, 2017. DOI: 10.1590/2317-1545v39n4177911. Acesso em: 12 jan. 2021.

LIMA, F. B. de. *et al.* F. Accelerated aging in sorghum genotypes. **Bioscience Journal**, Uberlândia, v. 35, n. 2, p. 450-458, 2019. DOI: 10.14393/BJ-v35n2a2019-41774. Acesso em: 12 jan. 2021.

LIN, H. *et al.* Qualitative and quantitative analysis of sorghum grain composition including protein and tannins using ATR-FTIR spectroscopy. **Food Analytical Methods**, [s.l.], v. 14, n. 2, p. 268-279, 2021. DOI: 10.1007/s12161-020-01874-5. Acesso em: 13 jul. 2022.

- MAGUIRE, J. D. Speed of germination-aid selection and evaluation for seedling emergence and vigor. **Crop Science**, [s.l.], v. 2, n. 1, p. 176-177, 1962. DOI: 10.2135/cropsci1962.0011183X000200020033x. Acesso em: 12 jan. 2021.
- MALINOWSKI, E. R. **Factor analysis in chemistry**. New York: J. Wiley, 1991. 350 p.
- MANUEL, M. N. B. *et al.* One-class classification of special agroforestry Brazilian coffee using NIR spectrometry and chemometric tools. **Food Chemistry**, [s.l.], v. 366, p. 130480, 2022. DOI: 10.1016/j.foodchem.2021.130480. Acesso em: 09 mar. 2022.
- MARCOS- FILHO, J. **Fisiologia de sementes de plantas cultivadas**. 2. ed. Londrina, PR: ABRATES, 2016. 660 p.
- MARCOS FILHO, J. Seed vigor testing: an overview of the past, present and future perspective. **Scientia Agricola**, Piracicaba, v. 72, n. 4, p. 363-374, 2015. DOI: 10.1590/0103-9016-2015-0007. Acesso em: 09 mar. 2022.
- MARTINS, L. dos S. *et al.* Adaptabilidade e estabilidade de híbridos de sorgo granífero no sudoeste de Goiás. **Revista Agrarian**, Dourados, v. 9, n. 34, p. 334-347, 2016. Disponível em: <https://www.embrapa.br/milho-e-sorgo/busca-de-publicacoes/-/publicacao/1023657/adaptabilidade-e-estabilidade-de-hibridos-de-sorgo-granifero-no-sudoeste-de-goias>. Acesso em: 17 jan. 2021.
- MAYR, S. *et al.* Quantification of Silymarin in Silybi mariani fructus: challenging the analytical performance of benchtop vs. handheld NIR spectrometers on whole seeds. **Planta Medica**, [s.l.], v. 88, n. 01, p. 20-32, 2021. DOI: 10.1055/a-1326-2497. Acesso em: 04 jul. 2022.
- MAYRINCK, L. G. *et al.* Use of near infrared spectroscopy in cotton seeds physiological quality evaluation. **Journal of Seed Science**, [s.l.], v. 42, p. 1-11, 2020. DOI: 10.1590/2317-1545v42227169. Acesso em: 05 nov. 2020.
- MBURU, M. *et al.* Spectroscopic analysis of chia seeds. **Scientific Reports**, [s.l.], v. 11, p. 9253, 2021. DOI: 10.1038/s41598-021-88545-5. Acesso em: 06 jul. 2022.
- MEDEIROS, A. D. *et al.* Machine Learning for Seed Quality Classification: An Advanced Approach Using Merger Data from FT-NIR Spectroscopy and X-ray Imaging. **Sensors**, [s.l.], v. 20, n. 15, p. 4319-4319, 2020. DOI: 10.3390/s20154319. Acesso em: 13 jan. 2021.
- MEDEIROS, A. D.; PEREIRA, M. D. SAPL<sup>®</sup>: a free software for determining the physiological potential in soybean seeds. **Pesquisa Agropecuária Tropical**, Goiânia, v. 48, n. 3, p. 222-228, 2018. DOI: 10.1590/1983-40632018v4852340. Acesso em: 25 jan. 2021.
- MEDEIROS, M. L. da S. *et al.* Assessment oil composition and species discrimination of Brassicas seeds based on hyperspectral imaging and portable near infrared (NIR) spectroscopy tools and chemometrics. **Journal of Food Composition and Analysis**, [s.l.], v. 107, p. 104403, 2022. DOI:10.1016/j.jfca.2022.104403. Acesso em: 01 jul. 2022.

- MEDEIROS, M. L. S. *et al.* Adequação do teste de lixiviação de potássio em sementes de *Moringa oleifera*. **Ciência Florestal**, Santa Maria, v. 29, n. 2, p. 941-949, 2019. DOI: DOI: 10.5902/1980509832534. Acesso em: 12 jan. 2021.
- MEENU, M. *et al.* Settings Open Access Article Rapid Determination of  $\beta$ -Glucan Content of Hulled and Naked Oats Using near Infrared Spectroscopy Combined with Chemometrics. **Foods**, [s.l.], v. 11, n. 43, p. 1-18, 2022. DOI: 10.3390/foods11010043. Acesso em: 24 mar. 2022.
- MOREIRA, F. R. da C. *et al.* Substituição parcial do milho por sorgo granífero na alimentação de suínos nas fases de creche, crescimento e terminação. **Revista Brasileira de Saúde e Produção Animal**, Salvador, v. 15, n. 1, p. 94-107, 2014. Acesso em: 15 jan. 2021.
- MUKASA, P. *et al.* Determination of the viability of retinispora (*Hinoki cypress*) seeds using shortwave infrared hyperspectral imaging spectroscopy. **Journal of Near Infrared Spectroscopy**, [s.l.], v. 28, n. 2, p. 70-80, 2020. DOI: 10.1177/0967033519898890. Acesso em: 03 nov. 2020.
- MUKASA, P. *et al.* Determination of viability of Retinispora (*Hinoki cypress*) seeds using FT-NIR spectroscopy. **Infrared Physics & Technology**, [s.l.], v. 98, p. 62-68, 2019. DOI: 10.1016/j.infrared.2019.02.008. Acesso em: 30 jun. 2022.
- MYINT, D. *et al.* Sustainable sesame (*Sesamum indicum* L.) production through improved technology: An overview of production, challenges, and opportunities in Myanmar. **Sustainability**, [s.l.], v. 12, n. 9, p. 3515, 2020. DOI: 10.3390/su12093515. Acesso em: 24 mar. 2022.
- NERY, M. C. *et al.* Accelerated ageing test and behaviour investigation of isoenzymes in sesame seeds. **Acta Scientiarum**, [s.l.], v. 40, n. 1, p. e39449, 2018. Disponível em: <https://doi.org/10.4025/actasciagron.v40i1.39449>. Acesso em: 24 mar. 2021.
- OLESEN, M. H. *et al.* Classification of Viable and Non-Viable Spinach (*Spinacia Oleracea* L.) Seeds by Single Seed near Infrared Spectroscopy and Extended Canonical Variates Analysis. **Journal of Near Infrared Spectroscopy**, [s.l.], v. 19, n. 3, p. 171-180, 2011. DOI: 10.1255/jnirs.928. Acesso em: 03 nov. 2020.
- OSGUERA-TOLEDO, M. E. *et al.* Physicochemical changes of starch during malting process of sorghum grain. **Journal of Cereal Science**, [s.l.], v. 95, p. 103069, 2020. DOI: 10.1016/j.jcs.2020.103069. Acesso em: 30 jun. 2022.
- PANTOJA, P. A. **Aplicação da espectroscopia de infravermelho próximo na caracterização da carga de petróleo para o processo de destilação atmosférica**. 2006. p. 91. Dissertação (Mestrado em Engenharia) - Universidade de São Paulo, São Paulo, 2006.
- PASQUINI, C. Near infrared spectroscopy: a mature analytical technique with new perspectives: a review. **Analytica Chimica Acta**, [s.l.], v. 14, n. 2, p. 198-219, 2003. Disponível em: <https://doi.org/10.1590/S0103-50532003000200006>. Acesso em: 03 mar. 2022.

- PASQUINI, C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. **Journal of the Brazilian Chemical Society**, [s.l.], v. 14, n. 2, p. 198-219, 2003. Disponível em: <https://doi.org/10.1590/S0103-50532003000200006>. Acesso em: 03 mar. 2022.
- PINTO, C. A. G. *et al.* Image analysis in the evaluation of the physiological potential of maize seeds. **Revista Ciência Agronômica**, Fortaleza, v. 46, n. 2, p. 319-328, 2015. DOI: 10.5935/1806-6690.20150011. Acesso em: 25 jan. 2021.
- QIU, G. *et al.* FT-NIR Spectroscopy for Detecting Supersweet Corn (*Zea mays* L. *Saccharata* Sturt) Seed Viability with Multivariate Data Analysis. **Sensors**, [s.l.], v. 18, n. 4, p. 1010, 2018. DOI:10.3390/s18041010. Acesso em: 05 jul. 2022.
- QUENTIN, A. G. *et al.* Application of near-infrared spectroscopy for estimation of non-structural carbohydrates in foliar samples of *Eucalyptus globulus* Labillardière. **Tree Physiology**, [s.l.], v. 37, p. 131-141, 2016. DOI: 10.1093/treephys/tpw083. Acesso em: 25 jan. 2021.
- RATAJCZAK, E. *et al.* The production, localization and spreading of reactive oxygen species contributes to the low vitality of long-term stored common beech (*Fagus sylvatica* L.) seeds. **Journal of Plant Physiology**, [s.l.], v. 174, p. 147-156, 2015. DOI: 10.1016/j.jplph.2014.08.021. Acesso em: 14 jan. 2021.
- REED, R. C.; BRADFORD, K. J.; KHANDAY, I. Seed germination and vigor: ensuring crop sustainability in a changing climate. **Heredity**, [s.l.], v. 128, n. 6, p. 450-459, 2022. DOI: 10.1038/s41437-022-00497-2. Acesso em: 12 jul. 2021.
- REICH, G. Near-infrared spectroscopy and imaging: basic principles and pharmaceutical applications. **Advanced Drug Delivery Reviews**, [s.l.], v. 57, n. 8, p. 1109-1143, 2005. DOI: 10.1016/j.addr.2005.01.020. Acesso em: 12 jul. 2021.
- RIBEIRO, J. P. O. *et al.* FT-NIR and linear discriminant analysis to classify chickpea seeds produced with harvest aid chemicals. **Food Chemistry**, [s.l.], v. 342, p. 128324, 2021. DOI:10.1016/j.foodchem.2020.128324. Acesso em: 06 jul. 2022.
- RUIZ-PEREZ, D. *et al.* So you think you can PLS-DA? **Bmc Bioinformatics**, [s.l.], v. 21, n. 1, p. 1-10, 2020. DOI: 10.1186/s12859-019-3310-7. Acesso em: 13 out. 2022.
- SAMAMAD, N. T. I. *et al.* Near infrared spectroscopy, a suitable tool for fast phenotyping – The case of cashew genetic improvement. **Scientia Horticulturae**, [s.l.], v. 238, p. 363-368, 2018. DOI: 10.1016/j.scienta.2018.05.007. Acesso em: 19 nov. 2020.
- SANTANA, F. *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**, [s.l.], v. 43, n. 3, p. 371-381, 2020. DOI: 10.21577/0100-4042.20170480. Acesso em: 02 ag. 2022.
- SARMENTO, E. C. S. *et al.* Physiological potential of sorghum seeds under discontinuous hydration and water deficiency conditions. **Revista Ciência Agronômica**, Fortaleza, v. 51, n. 4, e20207200, 2020. DOI: 10.5935/1806-6690.20200069. Acesso em: 25 jan. 2021.

SHARABY, N.; BUTOVCHENKO, A. Cultivation technology of sesame seeds and its production in the world and in Egypt. In: **IOP Conference Series: Earth and Environmental Science**, [s.l.], v. 403, n. 1, p. 1-9, 2019. DOI: 10.1088/1755-1315/403/1/012093. Acesso em: 03 mar. 2022.

SHI, H; YU, P. Comparison of grating-based near-infrared (NIR) and Fourier transform mid-infrared (ATR-FT/MIR) spectroscopy based on spectral preprocessing and wavelength selection for the determination of crude protein and moisture content in wheat. **Food Control**, [s.l.], v. 82, p. 57–65, 2017. DOI: 10.1016/j.foodcont.2017.06.015. Acesso em: 25 jan. 2021.

SIESLER, H. W. *et al.* **Near-infrared spectroscopy: principles, instruments, applications**. Weinheim: Wiley-VCH Verlage, 2002. 349 p.

SILVA, A. C. da *et al.* Salicylic acid as attenuator of drought stress on germination and initial development of sesame. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 21, n. 3, p. 156-162, 2017. Disponível em: <https://doi.org/10.1590/1807-1929/agriambi.v21n3p156-162>. Acesso em: 02 jan. 2020.

SILVA, A. G. da *et al.* Consórcio sorgo e braquiária na entrelinha para produção de grãos, forragem e palhada na entressafra. **Revista Ceres**, Viçosa, v. 61, n. 5, p. 697-705, 2014. DOI: 10.1590/0034-737X201461050013. Acesso em: 14 jan. 2021.

SILVA, R. T. da *et al.* Physiological quality of sesame seeds produced from plants subjected to water stress. **Revista Ciência Agronômica**, Fortaleza, v. 47, n. 4, p. 643-648, 2016. DOI: 10.5935/1806-6690.20160077. Acesso em: 25 jan. 2021.

SILVA, S. F. da. **Uso da espectrometria no infravermelho próximo na avaliação de qualidade de sementes de braquiária**. Dissertação (mestrado acadêmico) – Universidade Federal do Ceará, Fortaleza. 2019. 48 p

SILVA, V. N.; DOTTO, L. Análise de imagens para avaliação do crescimento de plântulas de arroz. **Agrarian Academy**, Goiânia, v. 4, n. 7, p. 39, 2017. DOI: 10.18677/Agrarian\_Academy\_2017a4. Acesso em: 18 jan. 2021.

SIMEONE, M. L. F.; RIBEIRO, M. R.; TRINDADE, R. dos S. **Espectroscopia no infravermelho próximo e análise discriminante por mínimos quadrados parciais como método alternativo para a seleção de sementes haploides de milho**. Sete Lagoas: Embrapa Milho e Sorgo, 2019. 19 p. (Boletim de Pesquisa e Desenvolvimento / Embrapa Milho e Sorgo, ISSN 1679-0154; 200). Acesso em: 11 fev. 2022.

SOUSA, E. C. *et al.* Classification of *Sorghum bicolor* (L.) Moench Seed Vigor by Respiratory Activity. **Journal of Agricultural Science**, [s.l.], v. 10, n. 10, p. 343-348, 2018. DOI: 10.5539/jas.v10n10p343. Acesso em: 18 jan. 2021.

VALERIANO, F. R. *et al.* Morfologia de sementes de gergelim. **Acta Iguazu**, Cascavel, v. 8, n. 2, p. 23-36, 2019. Disponível em: <<https://e-revista.unioeste.br/index.php/actaiguazu/article/view/19338/14388>>. Acesso em: 15 mar. 2022.

- VASCONCELOS, M. C. *et al.* Diferenciação de cultivares de girassol por espectroscopia no infravermelho próximo e análise multivariada, utilizando sementes e óleo. **Revista Brasileira de Ciências Agrárias**, Recife, v. 13, n. 4, p. 1-7, 2018. DOI: 10.5039/agraria.v13i4a5582. Acesso em: 07 fev. 2022.
- VIANA, L. C. *et al.* Modelos de calibração e a espectroscopia no infravermelho próximo para predição das propriedades químicas e da densidade básica da madeira de *Eucalyptus*. **Ciência Florestal**, Santa Maria, v. 20, n. 2, p. 367-376, 2010. Acesso em: 26 nov. 2021.
- WAKHOLI, C. *et al.* Rapid assessment of corn seed viability using short wave infrared line-scan hyperspectral imaging and chemometrics. **Sensors and Actuators B: Chemical**, [s.l.], v. 255, p. 498-507, 2018. DOI: 10.1016/j.snb.2017.08.036. Acesso em: 03 nov. 2020.
- WANG, Y. *et al.* Discriminant analysis and comparison of corn seed vigor based on multiband spectrum. **Computers and Electronics in Agriculture**, [s.l.], v. 190, p. 106444, 2021. DOI: 10.1016/j.compag.2021.106444. Acesso em: 04 jul. 2022.
- WANG, Y. *et al.* Feasibility analysis of NIR for detecting sweet corn seeds vigor. **Journal of Cereal Science**, [s.l.], v. 93, p. 102977, 2020. DOI: 10.1016/j.jcs.2020.102977. Acesso em: 29 jun. 2022.
- WILLIAMS, P.; NORRIS, K. **Near-Infrared Technology in the Agricultural and Food Industries**, 2. ed. American Association of Cereal Chemists Inc., St Paul, MN (2001).
- WORKMAN, JR. J.; WEYER, L. **Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy**. CRC Press. 2012. Boca Raton. 326.
- XAVIER, J. B. *et al.* Morphological, chemical and physiological characterization of *Amaranthus spp.* Seeds. **Journal of Seed Science**, [s.l.], v. 41, n. 4, p. 478-487, 2019. DOI: 10.1590/2317-1545v41n4226286. Acesso em: 15 mar. 2022.
- XU, Y.; ZHOU, Y.; CHEN, J. Near-Infrared Spectroscopy Combined with Multivariate Calibration to Predict the Yield of Sesame Oil Produced by Traditional Aqueous Extraction Process. **Journal of Food Quality**, [s.l.], v. 2017, p. 1-5, 2017. DOI: 10.1155/2017/2515476. Acesso em: 04 jul. 2022.
- YASMIN, J. *et al.* Classification Method for Viability Screening of Naturally Aged Watermelon Seeds Using FT-NIR Spectroscopy. **Sensors**, [s.l.], v. 19, n. 5, p. 1190, 2019. DOI: 10.3390/s19051190. Acesso em: 02 jul. 2022.

## APÊNDICE A – RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS E DOS PRÉ-TRATAMENTOS ESPECTRAIS EM SEMENTES DE GERGELIM

### Análise de componentes principais - ACP

Cargas fatoriais para o teste de germinação (G), primeira contagem de germinação (PC), índice de velocidade de germinação (IVG), emergência de plântulas (EP), índice de velocidade de emergência (IVE), condutividade elétrica (CE), massa seca de plântulas (MS), comprimento de plântulas (CP), envelhecimento acelerado de 24 (EA24h) e 48h (EA48h) em nove lotes de sementes de gergelim, autovalores, variância total e variância acumulada.

Variáveis	CP1	CP2
G	0,5141*	-0,0087359
PC	0,4255*	0,68317*
IVG	0,080738	0,040962
EM	0,50208*	0,048551
IVE	0,10712	0,18301
MS	0,000848992	0,0052197
CP	0,045028	0,12973
EA24h	0,38159	-0,64415
EA48h	0,37005	-0,25276
Variância total (%)	95,25	2,66
Variância acumulada (%)	95,25	97,91

\*Atributos com a maior contribuição em cada componente. CP = componente principal

### Calibração e validação – PLS-DA

Resultados estatísticos da calibração e validação interna dos modelos de PLS – DA em sementes de gergelim com alto (A), médio (B) e baixo vigor (C) respectivamente, submetidos a diferentes pré-tratamentos espectrais.

(A)

Pré-tratamento	Calibração (n=19)		Validação interna cruzada (n=19)		
	R <sup>2</sup>	RMSEC	R <sup>2</sup>	RMSECV	VL
Sem pré-tratamento	0,86	0,339	0,28	0,829	7
Smoothing	0,98	0,002	0,98	0,002	3
Smoothing + MSC	0,90	0,288	0,20	0,876	7
Smoothing + Primeira derivada	0,66	0,535	0,53	0,668	3
Smoothing + Segunda derivada	0,90	0,285	0,70	0,531	4
Smoothing + SNV	0,90	0,293	0,16	0,896	7
MSC	0,91	0,270	0,20	0,874	7
Primeira derivada Savitzky-Golay	0,67	0,529	0,54	0,663	3
Primeira derivada Savitzky-Golay + MSC	0,94	0,220	0,56	0,648	7
Primeira derivada Savitzky-Golay + SNV	0,94	0,220	0,55	0,653	7
Segunda derivada Savitzky-Golay	0,99	0,060	0,73	0,500	7
Segunda derivada Savitzky-Golay + MSC	0,89	0,305	0,64	0,581	2
Segunda derivada Savitzky-Golay + SNV	0,89	0,304	0,65	0,580	2
SNV	0,91	0,274	0,17	0,892	7

(B)

Pré-tratamento	Calibração (n=19)		Validação interna cruzada (n=19)		
	R <sup>2</sup>	RMSEC	R <sup>2</sup>	RMSECV	VL
Sem pré-tratamento	0,80	0,431	0,42	0,769	6
<i>Smoothing</i>	0,79	0,477	0,42	0,774	6
<i>Smoothing</i> + MSC	0,89	0,306	0,18	0,917	7
<i>Smoothing</i> + Primeira derivada	0,76	0,468	0,46	0,744	7
<i>Smoothing</i> + Segunda derivada	0,89	0,315	0,52	0,703	4
<i>Smoothing</i> + SNV	0,89	0,319	0,13	0,945	7
MSC	0,90	0,290	0,17	0,922	7
Primeira derivada Savitizky-Golay	0,78	0,447	0,39	0,795	6
Primeira derivada Savitizky-Golay + MSC	0,17	0,876	0,06	0,984	5
Primeira derivada Savitizky-Golay + SNV	0,17	0,876	0,06	0,984	1
Segunda derivada Savitizky-Golay	0,99	0,056	0,64	0,610	7
Segunda derivada Savitizky-Golay + MSC	0,99	0,079	0,65	0,596	6
Segunda derivada Savitizky-Golay + SNV	0,99	0,079	0,65	0,594	6
SNV	0,90	0,303	0,13	0,949	7

(C)

Pré-tratamento	Calibração (n=19)		Validação interna cruzada (n=19)		
	R <sup>2</sup>	RMSEC	R <sup>2</sup>	RMSECV	VL
Sem pré-tratamento	0,98	0,098	0,91	0,293	7
<i>Smoothing</i>	0,98	0,100	0,90	0,244	7
<i>Smoothing</i> + MSC	0,98	0,128	0,91	0,278	6
<i>Smoothing</i> + Primeira derivada	0,99	0,041	0,96	0,174	7
<i>Smoothing</i> + Segunda derivada	0,99	0,077	0,88	0,335	7
<i>Smoothing</i> + SNV	0,98	0,126	0,92	0,273	6
MSC	0,98	0,125	0,91	0,277	6
Primeira derivada Savitizky-Golay	0,99	0,041	0,96	0,176	7
Primeira derivada Savitizky-Golay + MSC	0,99	0,054	0,94	0,231	7
Primeira derivada Savitizky-Golay + SNV	0,99	0,053	0,94	0,227	7
Segunda derivada Savitizky-Golay	0,99	0,083	0,85	0,371	6
Segunda derivada Savitizky-Golay + MSC	0,99	0,060	0,87	0,347	5
Segunda derivada Savitizky-Golay + SNV	0,99	0,060	0,87	0,347	5
SNV	0,98	0,123	0,92	0,273	6

R<sup>2</sup> – Coeficiente de determinação; RMSECV - Raiz quadrado do erro médio para validação cruzada; RMSEP - Raiz quadrado do erro médio para predição; VL – Variáveis latentes; n- número de amostras.

## APÊNDICE B – RESULTADOS DOS PRÉ-TRATAMENTOS ESPECTRAIS EM SEMENTES DE SORGO

Resultados estatísticos da calibração e validação interna dos modelos de PLS – DA em sementes de sorgo de diferentes níveis de vigor, submetidos a diferentes pré-tratamentos espectrais.

Pré-tratamento	Calibração (n=8)		Validação cruzada (n=8)		
	R <sup>2</sup>	RMSEC	R <sup>2</sup>	RMSECV	VL
Sem pré-tratamento	0,99	0,0495	0,83	0,3967	4
<i>Smoothing</i>	0,99	0,0497	0,83	0,3963	4
<i>Smoothing</i> + MSC	0,62	0,5995	0,57	0,6488	1
<i>Smoothing</i> + Primeira derivada	0,63	0,5225	0,54	0,6689	1
<i>Smoothing</i> + Segunda derivada	0,61	0,5344	0,45	0,7498	1
<i>Smoothing</i> + SNV	0,63	0,5219	0,59	0,6319	1
MSC	0,62	0,5294	0,57	0,6487	1
MSC + Primeira derivada	0,65	0,5077	0,58	0,6372	1
MSC + Segunda derivada	0,67	0,4907	0,69	0,6213	1
MSC + SNV	0,63	0,5218	0,59	0,6318	1
Primeira derivada Savitzky-Golay	0,63	0,5224	0,54	0,6682	1
Segunda derivada Savitzky-Golay	0,63	0,526	0,44	0,7399	1
SNV	0,63	0,5218	0,59	0,6318	1
SNV + Primeira derivada	0,66	0,5028	0,59	0,6326	1
SNV + Segunda derivada	0,69	0,4785	0,59	0,6331	1

R<sup>2</sup> – Coeficiente de determinação; RMSECV - Raiz quadrado do erro médio para validação cruzada; RMSEP - Raiz quadrado do erro médio para predição; VL – Variáveis latentes; n- número de amostras.