

ANUNCIENE BARBOSA DUARTE

**ESTRATÉGIAS E SELEÇÃO DE CULTIVARES DE SOJA EM ESTÁDIOS INICIAIS
PARA TOLERÂNCIA AO DÉFICIT HÍDRICO**

Tese submetida à Universidade Federal de Viçosa
como parte das exigências do Programa de Pós-
Graduação em Fitotecnia para obtenção do título
de *Doctor Scientiae*.

Orientador: Felipe Lopes da Silva

**VIÇOSA - MINAS GERAIS
2023**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

D812e
2023 Duarte, Anunciene Barbosa, 1992-
Estratégias e seleção de cultivares de soja em estádios
iniciais para tolerância ao déficit hídrico / Anunciene Barbosa
Duarte. – Viçosa, MG, 2023.
1 tese eletrônica (62 f.): il. (algumas color.).

Orientador: Felipe Lopes da Silva.
Tese (doutorado) - Universidade Federal de Viçosa,
Departamento de Agronomia, 2023.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2023.231>

Modo de acesso: World Wide Web.

1. Soja - Resistência à seca. 2. Aprendizado do computador.
3. Análise espectral. I. Silva, Felipe Lopes da, 1981-.
II. Universidade Federal de Viçosa. Departamento de
Agronomia. Programa de Pós-Graduação em Fitotecnia.
III. Título.

CDD 22. ed. 632.12

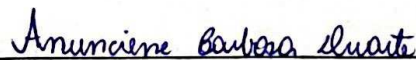
ANUNCIENE BARBOSA DUARTE

**ESTRATÉGIAS E SELEÇÃO DE CULTIVARES DE SOJA EM ESTÁDIOS INICIAIS
PARA TOLERÂNCIA AO DÉFICIT HÍDRICO**

Tese submetida à Universidade Federal de Viçosa
como parte das exigências do Programa de Pós-
Graduação em Fitotecnia para obtenção do título
de *Doctor Scientiae*.

APROVADA: 17 de fevereiro de 2023.

Assentimento:



Anunciene Barbosa Duarte
Autora



Felipe Lopes da Silva
Orientador

AGRADECIMENTOS

A Deus, por me conceder mais essa graça. A Ele toda honra, glória, louvor e exaltação.

À Nossa Senhora, que esteve sempre comigo zelando e intercedendo por mim e pela minha família. Ela cuidou de todos os detalhes e providenciou para que não nos faltasse o “melhor vinho”.

Aos meus pais, a quem agradeço principalmente por me concederem a vida e me ensinar valores e princípios que foram essenciais para que eu ingressasse nessa jornada.

Ao meu esposo, agradeço pelo amor, suporte e por estar sempre perto em todos os momentos.

Ao meu pequeno grande homem, Samuel, por me permitir receber de Deus o maior e mais nobre de todos os títulos. Com ele tenho a oportunidade de aprender a ser mãe a cada amanhecer.

Aos meus irmãos, pelo carinho, conselhos e torcida. Sem eles eu não teria conseguido.

Aos cunhados, pelos momentos descontraídos que tornaram essa caminhada mais leve.

À Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Fitotecnia pela oportunidade de cursar o doutorado.

Ao meu orientador, o Professor Felipe Lopes, pelo apoio, compreensão e por todos os ensinamentos transmitidos.

Aos membros que compuseram a banca de defesa desta tese (Felipe Lopes, Francisco Charles, Lorena Moura, Maicon Nardino e Martha Freire) agradeço a disponibilidade e valiosas contribuições para a melhoria deste material.

Às minhas orientadoras de iniciação científica, Silvia Nietzsche e Márcia Costa, com quem tive o prazer de conviver e aprender muito durante minha graduação.

Aos amigos do MUR, que foram sinais de Deus para mim em Viçosa.

Aos amigos do laboratório de soja/UFV, especialmente ao companheiro de experimento Dalton, pela ajuda, risadas, parceria e por todos os momentos compartilhados.

Aos amigos que Viçosa me deu, os quais terão sempre um cantinho muito especial no meu coração, minha sincera gratidão por cada momento vivido. Foi muito bom dividir a vida com vocês durante esses anos.

As agências financiadoras, pelo suporte financeiro para esta pesquisa se concluísse.

O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), da Fundação de Amparo à Pesquisa do Estado de Minas

Gerais (FAPEMIG) e da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

RESUMO

DUARTE, Anunciene Barbosa, D.Sc., Universidade Federal de Viçosa, fevereiro de 2023. **Estratégias e seleção de cultivares de soja em estádios iniciais para tolerância ao déficit hídrico.** Orientador: Felipe Lopes da Silva.

A soja é uma importante oleaginosa em todo o mundo e apresenta sensibilidade ao déficit hídrico, especialmente na fase inicial de desenvolvimento. Por isso, avaliar diferentes estratégias buscando selecionar cultivares de soja no início do desenvolvimento é extremamente necessário. Nesse sentido, são objetivos deste estudo: i) selecionar cultivares de soja tolerantes ao déficit hídrico utilizando os índices de seleção FAI-BLUP (Factor analysis and ideotype-design - Best Linear Unbiased Prediction) e o MGIDI (multi-trait genotype–ideotype distance index), além de indicar cultivares tolerantes ao déficit hídrico na fase inicial de desenvolvimento; ii) prever o conteúdo de água nas folhas de soja e classificá-las quanto a condição hídrica, por meio de dados de espectroscopia NIR (near infrared) e diferentes modelos de machine learning. Dois experimentos foram avaliados envolvendo duas formas de imposição de estresse por déficit hídrico (em solo e em areia) em dois estádios diferentes (germinação e V1) e repetidos por duas épocas. Os experimentos envolveram 100 cultivares de soja, as quais foram submetidas a duas condições de disponibilidade hídrica (condição controle e condição estresse). Em ambos os experimentos, o estresse permaneceu por 20 dias. Os índices de seleção FAI-BLUP e MGIDI possibilitaram a seleção de 15 cultivares de soja. Foram selecionadas 12 cultivares comuns aos dois índices. As cultivares M 9144 RR, BMX TITAN RR foram as que mais se aproximaram do ideótipo. Em relação aos modelos de machine learning, todos os quatro modelos utilizados apresentaram boas performances ao realizar as tarefas de classificação e regressão. Os modelos PLS (Partial Least Squares) e SVM (Support Vector Machine) apresentaram os melhores resultados para classificar folhas de soja quanto à condição hídrica. Já para a tarefa de regressão, os modelos PLS e PCR (principal component regression) apresentaram os melhores desempenhos.

Palavras-chave: Índices de seleção. Machine learning. Espectroscopia NIR. Seca.

ABSTRACT

DUARTE, Anunciene Barbosa, D.Sc., Universidade Federal de Viçosa, February, 2023. **Strategies for early selection of soybean cultivars for tolerance to water deficit.** Adviser: Felipe Lopes da Silva.

Soybean is an important oilseed in Brazil and in the world and is sensitive to water deficit, especially in the initial stage of development. For this reason, this thesis evaluates different strategies seeking to mitigate the effects of water deficit in the soybean crop in the initial stage of development. In this sense, the objectives of this study are: i) to select soybean cultivars tolerant to water deficit using the selection indices FAI-BLUP (Factor analysis and ideotype-design - Best Linear Unbiased Prediction) and MGIDI (multi-trait genotype–ideotype distance index), in addition to indicate cultivars tolerant to water deficit in the initial stage of development; ii) predict the water content in soybean leaves and classify them according to their water condition using NIR (near infrared) spectroscopy data and different machine learning models. Two experiments were carried out involving two forms of water deficit stress imposition (in soil and in sand) in two different stages (germination and V1) and they were repeated for two seasons. The experiments involved 100 soybean cultivars, which were subjected to two conditions of water availability (control condition and stress condition). In both experiments, the stress remained for 20 days. The selection indices FAI-BLUP and MGIDI allowed the selection of 15 soybean cultivars. Twelve cultivars common to both indices were selected. Cultivars M 9144 RR, BMX TITAN RR were the closest to the ideotype. Regarding the machine learning models, all the four models showed good results when performing the classification and regression tasks. The PLS (Partial Least Squares) and SVM (Support Vector Machine) models presented the best results to classify soybean leaves regarding their hydric condition. For the regression task, the PLS and PCR (principal component regression) models showed the best performances.

Keywords: Selection indices. Machine learning. NIR spectroscopy. Drought.

SUMÁRIO

INTRODUÇÃO GERAL	8
REFERÊNCIAS	10
Artigos científicos.....	12
Seleção de cultivares de soja tolerantes ao deficit hídrico na fase inicial de desenvolvimento	13
1. INTRODUÇÃO.....	14
2. MATERIAL E MÉTODOS.....	16
Experimento 1. Imposição de deficit hídrico na germinação de sementes e desenvolvimento inicial de plântulas de soja.....	17
Experimento 2. Imposição de deficit hídrico no estágio fenológico v1	19
3. RESULTADOS E DISCUSSÃO.....	24
4. CONCLUSÕES	34
5. REFERÊNCIAS	34
Espectroscopia NIR para predição do conteúdo de água em folhas de soja e classificação quanto à disponibilidade hídrica	40
1. INTRODUÇÃO.....	41
2.MATERIAL E MÉTODOS.....	43
3. RESULTADOS E DISCUSSÃO.....	47
4. CONCLUSÕES	57
5. REFERÊNCIAS	58
CONCLUSÕES GERAIS.....	62

INTRODUÇÃO GERAL

A soja é uma importante oleaginosa em todo o mundo, sendo rica em vitaminas, óleos, e uma excelente fonte de proteínas, além de ser a principal matéria prima para a produção de biocombustível (Chinakwe et al., 2019). O Brasil assume lugar de destaque na produção mundial de soja com produção em torno de 130 milhões de toneladas (Conab, 2021). Entretanto, apesar do crescente incremento produtivo, cada vez mais a cultura tem sofrido com algum tipo de estresse abiótico, reduzindo, portanto, o potencial produtivo da soja.

Dentre os estresses abióticos, o deficit hídrico é um dos mais preocupantes (Youssef et al., 2021). Dada a sensibilidade da soja, este tipo de estresse acarreta diversos prejuízos para a cultura refletindo na redução da produção (Liu et al., 2017; Seepaul et al., 2019; Ahanger et al., 2021). Tal redução varia de acordo com a duração e intensidade do estresse (Nemeskéri; Helyes, 2019; Feng et al., 2020).

Na tentativa de compreender o comportamento das plantas de soja frente ao deficit hídrico, muitos estudos já foram realizados envolvendo as mais diferentes áreas da agronomia (Mondani et al., 2019; Braga-reis et al., 2021; Youssef et al., 2021; Purdom et al., 2022). A maioria dos estudos tem sido focados na imposição do estresse por deficit hídrico na fase reprodutiva da cultura. Entretanto, é evidente que o deficit hídrico afeta o desenvolvimento da soja desde os estádios iniciais (Wang e Komatsu 2017) trazendo sérios prejuízos para a lavoura.

Conhecer a demanda hídrica da cultura da cultura é primordial para o bom manejo da lavoura. Durante o ciclo da soja, é necessário um volume acumulado variando de 450 a 800 mm de água para garantir elevada produtividade (Pas, 2005). De acordo com Zhang et al. (2022) a fase de plântula é a base do crescimento das plantas de soja, e o desenvolvimento da soja na fase de plântula determina a qualidade do desenvolvimento individual e populacional das plantas em uma lavoura (Hua et al., 2018). Vale ressaltar que o bom desenvolvimento inicial da planta irá refletir em altas produtividades de grãos na colheita (Liu et al., 2017). Por isso, é fundamental que os programas de melhoramento de soja visando a tolerância ao deficit hídrico busquem estudar os efeitos desse tipo de estresse na fase inicial do desenvolvimento, utilizando diferentes estratégias para seleção de cultivares tolerantes.

Na atualidade já existem muitas ferramentas que podem ser utilizadas nos programas de melhoramento envolvendo a seleção de cultivares em condições de estresse por deficit hídrico. Por exemplo, tem-se diferentes técnicas de modelagens, uso de espectroscopia NIR, bem como

uso de diferentes índices de seleção. Estas abordagens contribuem para que os melhoristas tomem decisões cada vez mais assertivas.

O uso dos índices de seleção é uma estratégia particularmente vantajosa, já que permite a seleção de várias características de interesse agrônômico de forma simultânea, contribuindo para o sucesso dos programas de melhoramento (Leite et al., 2018; Rangel et al., 2011; Woyann et al., 2020). Índices de seleção como o FAI-BLUP (Factor analysis and ideotype-design), proposto por Rocha et al. (2018) e o MGIDI (Multitrait Genotype-Ideotype Distance Index) proposto por Olivoto e Nardino (2021) tem ganhado espaço nas pesquisas científicas (Silva et al., 2018; Woyann et al., 2019; Pathy et al., 2021; Pessoa et al., 2021; Pour-Aboughadareh et al., 2021; Ahmed et al., 2022; Kistner et al., 2022; Nardino et al., 2022; Silva Junior et al., 2022). Estes índices, além de realizar a seleção para múltiplas características de forma simultânea, lida bem com problemas relacionados à multicolinearidade.

A espectroscopia de infravermelho próximo (NIR) é uma ferramenta poderosa para monitorar vários processos (Zhang et al., 2017) constituindo uma excelente aliada em estudos de restrição hídrica. De acordo com (Pasquini, 2018), o comportamento espectral de uma amostra constitui uma resposta às suas propriedades, criando, assim, uma assinatura espectral única de cada amostra. Dessa forma, esta abordagem constitui uma ferramenta promissora para determinar o teor de água foliar de forma mais rápida e eficiente e contribuir com os programas de melhoramento de soja para a seleção de cultivares de sojas tolerantes ao déficit hídrico.

Dado o exposto, o presente estudo apresenta os seguintes objetivos: i) selecionar cultivares de soja tolerantes ao déficit hídrico na fase inicial de desenvolvimento utilizando os índices de seleção FAI-BLUP e MGIDI; ii) prever o conteúdo de água nas folhas de soja e classificá-las quanto a condição hídrica por meio de dados NIR e diferentes modelos de machine learning.

REFERÊNCIAS

- Ahanger, M. A., Siddique, K. H., & Ahmad, P. (2021). Understanding drought tolerance in plants. *Physiol Plant*, *172*, 286-8.
- Ahmed, I., Rohman, M., Hossain, M., Molla, M., Azam, M., Hasan, M., ... & Hossain, A. (2022). A Study on the Phenotypic Variation of 103 Cucumber (*Cucumis sativus* L.) Landraces for the Development of Desirable Cultivars Suitable for the Changing Climate. *Life*, *12*(8), 1235.
- Braga-Reis, I., Neris, D. M., Ribas, A. F., Vieira, L. G. E., & Souza, G. M. (2021). Gamma-aminobutyric acid (GABA) and acetylcholine (ACh) alleviate water deficit effects in soybean: From gene expression up to growth performance. *Environmental and Experimental Botany*, *182*, 104303.
- Chinakwe, E. C., Ibekwe, V. I., Nwogwugwu, U. N., Ofoegbu, J., Mike-Anosike, E., Nwachukwu, I. N., ... & Chinakwe, P. O. (2019). Evaluation of plant growth promoting potentials exhibited by rhizobacteria associated with beans plant. *Malaysian Journal of Sustainable Agriculture (MJSA)*, *3*(1), 20-22.
- Feng, Z., Ding, C., Li, W., Wang, D., & Cui, D. (2020). Applications of metabolomics in the research of soybean plant under abiotic stress. *Food Chemistry*, *310*, 125914.
- Gupta, A., Rico-Medina, A., & Caño-Delgado, A. I. (2020). The physiology of plant responses to drought. *Science*, *368*(6488), 266-269.
- Leite, W. D. S., Unêda-Trevisoli, S. H., Silva, F. M. D., Silva, A. J. D., & Mauro, A. O. D. (2018). Identification of superior genotypes and soybean traits by multivariate analysis and selection index. *Revista Ciência Agronômica*, *49*, 491-500.
- Mondani, F., Khani, K., Honarmand, S. J., & Saeidi, M. (2019). Evaluating effects of plant growth-promoting rhizobacteria on the radiation use efficiency and yield of soybean (*Glycine max*) under water deficit stress condition. *Agricultural Water Management*, *213*, 707-713.
- Nemeskéri, E., & Helyes, L. (2019). Physiological responses of selected vegetable crop species to water stress. *Agronomy* *9*, 447.
- Olivoto, T., & Nardino, M. (2021). MGIDI: toward an effective multivariate selection in biological experiments. *Bioinformatics*, *37*(10), 1383-1389.
- Pas, C. (2005). Manual de segurança e qualidade para cultura da soja. Brasília: Embrapa.
- Pasquini, C. (2018). Near infrared spectroscopy: A mature analytical technique with new perspectives—A review. *Analytica chimica acta*, *1026*, 8-36.

- Pathy, T. L., & Mohanraj, K. (2021). Estimating best linear unbiased predictions (BLUP) for yield and quality traits in sugarcane. *Sugar Tech*, 23(6), 1295-1305.
- Pessoa, H. P., Da Silva, L. J., Chagas, R. R., Copati, M. G. F., Dariva, F. D., de Oliveira Dias, F., & Gomes, C. N. (2021). Uncovering tomato candidate genes for seed germination performance and seedling vigour using the *Solanum pennellii* introgression line population. *Scientia Horticulturae*, 284, 110134.
- Pour-Aboughadareh, A., Sanjani, S., Nikkhah-Chamanabad, H., Mehrvar, M. R., Asadi, A., & Amini, A. (2021). Identification of salt-tolerant barley genotypes using multiple-traits index and yield performance at the early growth and maturity stages. *Bulletin of the National Research Centre*, 45(1), 1-16.
- Purdom, S., Shekoofa, A., McClure, A., Pantalone, V., Arelli, P., & Duncan, L. (2022). Variation in mid-south soybean genotypes for recovery of transpiration rate and leaf maintenance following severe water-deficit stress. *Field Crops Research*, 286, 108625.
- Rangel, R. M., Amaral Júnior, A. T. D., Gonçalves, L. S. A., Freitas Júnior, S. D. P., & Candido, L. S. (2011). Análise biométrica de ganhos por seleção em população de milho pipoca de quinto ciclo de seleção recorrente. *Revista Ciência Agronômica*, 42, 473-481.
- Rocha, J. R. D. A. S. D. C., Machado, J. C., & Carneiro, P. C. S. (2018). Multitrait index based on factor analysis and ideotype-design: Proposal and application on elephant grass breeding for bioenergy. *Gcb Bioenergy*, 10(1), 52-60.
- Seepaul, D. D. R., George, S., Groot, J., & Wright, D. (2019). Drought tolerance classification of common oilseed species using seed germination assay. *Journal of Oilseed Brassica*, 10(2), 97-105.
- Youssef, M. S., Renault, S., Hill, R. D., & Stasolla, C. (2021). The soybean Phytoalbumin1 (GmPgb1) is involved in water deficit responses through changes in ABA metabolism. *Journal of Plant Physiology*, 267, 153538.
- Zhang, Y., Liu, Z., Wang, X., Li, Y., Li, Y., Gou, Z., ... & Qiu, L. (2022). Identification of genes for drought resistance and prediction of gene candidates in soybean seedlings based on linkage and association mapping. *The Crop Journal*, 10(3), 830-839.

Artigos científicos

Artigo 1: Seleção de cultivares de soja tolerantes ao deficit hídrico na fase inicial de desenvolvimento

Seleção de cultivares de soja tolerantes ao deficit hídrico na fase inicial de desenvolvimento

Resumo: Dada a sensibilidade da soja ao deficit hídrico, a busca por materiais tolerantes a este tipo de estresse continua sendo de grande interesse para os programas de melhoramento. Nesse sentido, os objetivos do presente estudo foram: selecionar cultivares de soja tolerantes ao deficit hídrico utilizando os índices de seleção FAI-BLUP e MGIDI além de indicar cultivares tolerantes ao deficit hídrico na fase inicial de desenvolvimento. Dois experimentos foram conduzidos em duas épocas em casa de vegetação. Ambos os experimentos envolveram 100 cultivares de soja, as quais foram submetidas a duas condições de disponibilidade hídrica (condição controle e condição estresse (deficit hídrico)). No primeiro experimento, a condição controle constituiu no uso de substrato sem restrição hídrica (areia umedecida com água), enquanto na condição estresse, havia deficit hídrico, o qual foi simulado utilizando PEG 6000 (-0.2 mPa). O estresse foi imposto na germinação. Já no segundo experimento, na condição controle o solo foi mantido na capacidade de campo (-33 kPa) e na condição estresse o solo foi mantido sob uma tensão de 1200 kPa. Neste caso, o estresse foi imposto quando as plantas de soja atingiram o estágio fenológico v1. Em ambos os experimentos, o estresse permaneceu por 20 dias. Foram utilizados os índices de seleção FAI-BUP e MGIDI para realizar o ranqueamento das cultivares com potencial de tolerância ao deficit hídrico. Por meio da análise de fatores, as variáveis utilizadas neste estudo foram agrupadas em quatro fatores. A variância acumulada dos primeiros quatro componentes principais foi responsável por aproximadamente 91% de toda a variância presente no conjunto de dados. Os índices de seleção FAI-BLUP e MGIDI possibilitaram a seleção de 15 cultivares de soja. Foram selecionadas 12 cultivares comuns aos dois índices. As cultivares M 9144 RR, BMX TITAN RR foram as que mais se aproximaram do ideótipo. As cultivares de soja selecionadas no presente estudo podem ser utilizadas como fonte de germoplasma em estudos futuros em programas de melhoramento visando a tolerância ao deficit hídrico.

Palavras-chave: Seca; Índice de seleção; FAI-BLUP; MGIDI.

1. INTRODUÇÃO

A maior parte das terras cultiváveis frequentemente são acometidas por algum tipo de estresse abiótico. O estresse ocasionado pelo déficit hídrico assume destaque, visto as graves consequências ao longo do ciclo das culturas, ameaçando o sucesso da produção agrícola (Bhatia et al., 2016; Ahanger et al., 2021; Zhang et al., 2022) e impedindo as plantas de atingirem seu máximo potencial genético.

A cultura da soja é particularmente sensível ao déficit hídrico em todos os estádios de desenvolvimento (Seepaul et al., 2019; Wang et al., 2022; Zhang et al., 2022). Quando este estresse acomete a lavoura, ocorre redução da germinação, atraso no estabelecimento da lavoura, redução de crescimento (Wang et al., 2022), atraso no ciclo da cultura, além de contribuir para redução da qualidade de semente (Vurukonda et al., 2016) e da produção (He et al., 2017; Martins et al., 2019; Wang et al. 2022).

O déficit hídrico é especialmente preocupante para a fase de germinação das sementes e fases iniciais do desenvolvimento. Isso porque a restrição hídrica afeta o processo de embebição da semente durante o processo germinativo. Em consequência disso, tem-se vários distúrbios metabólicos (Bewley e Black 2012), além de redução significativa na porcentagem e velocidade de germinação (Hameed et al., 2021).

Dessa forma, o déficit hídrico na fase inicial de desenvolvimento da soja pode gerar graves consequências ao longo do ciclo da cultura, incluindo redução da produção (Liu et al., 2017). A fase de plântula é a base do crescimento das plantas de soja (Zhang et al., 2022). Dessa forma, o desenvolvimento da soja na fase de plântula determina a qualidade do desenvolvimento individual e populacional das plantas em uma lavoura (Hua et al., 2018). Por isso, é cada vez mais oportuno estudar os efeitos do déficit hídrico na fase inicial do desenvolvimento de diferentes cultivares de soja.

Nesse sentido, um dos grandes esforços dos programas de melhoramento de soja visando a tolerância ao déficit hídrico tem sido a incansável busca por materiais tolerantes, que lidem bem com este tipo de adversidade ambiental, sem acarretar perdas no rendimento da lavoura (Fleury et al., 2010). A constante busca por materiais tolerantes envolve a avaliação de múltiplas características (Santchurn et al., 2014). Assim, para realizar a tarefa de seleção, comumente os melhoristas lançam mão de diferentes ferramentas estatísticas, como os índices de seleção.

O uso dos índices de seleção é uma estratégia vantajosa, já que permite a seleção de várias características de interesse agrônomo de forma simultânea, contribuindo para o sucesso

dos programas de melhoramento (Leite et al., 2018; Rangel et al., 2011; Woyann et al., 2020). O primeiro índice de seleção utilizado no melhoramento vegetal capaz de realizar a seleção simultânea, foi proposto por Smith-Hazel (Smith, 1936; Hazel, 1943) e ainda tem sido muito usado atualmente. Apesar do sucesso deste índice, ele apresenta limitações que reduz a precisão no processo de seleção, levando a conclusões errôneas (Woyann et al., 2019).

Nesse sentido, dada as deficiências dos índices de seleção mais clássicos, outros índices foram propostos com o passar dos anos para auxiliar os melhoristas na escolha de materiais superiores (Ceron-Rojas e Crossa 2018; Zuffo et al., 2020) com ganhos genéticos satisfatórios (Bermudez e Pinheiro, 2020). Índices de seleção como o FAI-BLUP (Factor analysis and ideotype-design), proposto por Rocha et al. (2018) e o MGIDI (Multitrait Genotype-Ideotype Distance Index) proposto por Olivoto et al. (2021) têm ganhado espaço nas pesquisas científicas. Estes índices, além de realizar a seleção para múltiplas características de forma simultânea, não dependem de pesos econômicos e lidam bem com problemas relacionados à multicolinearidade.

Em resumo, o índice FAI-BLUP leva em consideração a estrutura da correlação obtida a partir dos dados e direciona para selecionar genótipos mais próximos daquele hipotetizado pelo melhorista (ideótipo) (Rocha et al., 2018). Este índice de seleção utiliza a análise de fatores de forma a extrair as cargas fatoriais da matriz de correlação genética. Posteriormente, é realizado a estimação da distância ideótipo-genótipo, possibilitando o ranqueamento dos genótipos. Muitos trabalhos têm sido realizados envolvendo este índice para seleção de genótipos em diversas culturas, como tomate (Pessoa et al., 2021), Soja (Woyann et al., 2019), Sorgo (Silva et al., 2018), milho (Kistner et al., 2022) e arroz (Silva Junior et al., 2022).

O índice MGIDI, se baseia no redimensionamento das características, utilização de análise de fatores para explicar a estrutura de correlação e redução de dimensionalidade dos dados, além do planejamento de um ideótipo com base em valores conhecidos e o cálculo da distância entre cada genótipo e o ideótipo. Também este índice foi utilizado para seleção de genótipos de pepino (Ahmed et al., 2022), trigo (Nardino et al., 2022), cevada (Pour-Aboughadareh et al., 2021), cana de açúcar (Pathy et al., 2021) dentre outras culturas.

No caso da soja, alguns estudos já foram realizados visando o entendimento da tolerância ao deficit hídrico no estágio reprodutivo (Mengistu et al. 2018; Du et al., 2020; Khatun et al., 2021; Marinho et al., 2022; Rasheed et al., 2022). Entretanto, estudos que priorizem seleção de cultivares tolerantes a este estresse nos estádios iniciais de desenvolvimento ainda são escassos na literatura. Dado o exposto, os objetivos deste estudo

foram: i) selecionar cultivares de soja tolerantes ao deficit hídrico utilizando os índices de seleção FAI-BLUP e MGIDI; ii) indicar cultivares tolerantes ao deficit hídrico na fase inicial de desenvolvimento.

2. MATERIAL E MÉTODOS

Material Vegetal

Dois experimentos foram conduzidos em casa de vegetação envolvendo 100 cultivares de soja, pertencentes ao banco de germoplasma da Universidade Federal de Viçosa (Tabela 1).

Tabela 1. Descrição das 100 cultivares utilizadas neste estudo e suas respectivas identificações (id).

Cultivar	Id.	Cultivar	Id.	Cultivar	Id.
NA 4990 RG	1	BRS 232	35	IAS 5	69
CD 216	2	BRS 283	36	CD/FAPA 220	70
CD 215	3	5D 711 RR	37	R7	71
BMX ENERGIA RR	4	BRS 243 RR	38	FT GUAÍRA	72
BMX TITAN RR	5	FUNDACEP 57 RR	39	MGBR 48 Garimpo	73
NS 4823	6	5D 690 RR	40	EMBRAPA 59	74
BMX TURBO	7	BMX POTENCIA	41	5G770 RR	75
SYN 3358 RR	8	BRS 246 RR	42	BRS 213	76
CD 250 RR	9	5D 660 RR	43	FUNDACEP 33	77
A 4725 RG	10	FUNDACEP 58 RR	44	CD254RR	78
CD 225RR	11	FUNDACEP 53 RR	45	RA 518 RR	79
M-SOY 6101	12	CD 231 RR	46	Ocepar 14	80
BRAGG	13	CD 232	47	MG/BR Conquista	81
CD 5807	14	M 7211 RR	48	CD 2840	82
EMGOPA 304 (Campeira)	15	BRS 268	49	CD 2800	83
BRSMG RENASCENÇA	16	TMG 1066 RR	50	P98Y51	84
OCEPAR 3 PRIMAVERA	17	M 6707 RR	51	P98Y70	85
A 8000	18	BMX MAGNA	52	CD 246	86
MERCEDES 70 A	19	CD 226RR	53	CD 237 RR	87
A 6001 RR	20	CD 201	54	P98Y11	88
CD 224	21	CD 242RR	55	CD 243 RR	89
BRS 245 RR	22	CD 240RR	56	CD 244 RR	90
CD 208	23	NK 7059 RR	57	M-SOY 7901	91
BRS 282	24	BRS 284	58	TMG115RR	92
BRS 185	25	CD 236 RR	59	M-SOY 8001	93
BRS 133	26	CD 2630 RR	60	M9144RR	94
CD 206	27	CD 202	61	CD 234RR	95
TMG 1067 RR	28	RA 628 RR	62	CD 205	96
EMBRAPA 48	29	CD 248 RR	63	CD 2792 RR	97
CD 217	30	CD 235 RR	64	M 6009 RR	98
BRS 257	31	NA 5909 RG	65	CD 253	99
CD 229 RR	32	CD 233 RR	66	BRSMT (PINTATO)	100
FUNDACEP 56 RR	33	CD 2721RR	67		
ANTA 82	34	CD 2737 RR	68		

Ambos os experimentos foram conduzidos sob o delineamento de blocos casualizados em esquema fatorial envolvendo 2 fatores com três repetições. O fator 1 (cultivar) foi definido por 100 cultivares de soja e o fator 2 (disponibilidade hídrica), por 2 condições de disponibilidade hídrica, controle e estresse (deficit hídrico). A fim de obter maior confiabilidade nos dados coletados, os dois experimentos foram realizados em duas épocas diferentes (nos meses de abril e maio para o experimento 1 e nos anos de 2019 e 2021 para o experimento 2).

Experimento 1. Imposição de deficit hídrico na germinação de sementes e desenvolvimento inicial de plântulas de soja

O experimento foi instalado em casa de vegetação na Universidade Federal de Viçosa. Foram utilizadas sementes de 100 cultivares de soja, as quais foram tratadas com Cropstar (tiodicarbe e imidacloprido) + Maxim quatro (azoxistrobina, tiabendazol, fludioxonil e Metalaxil-M) e semeadas em bandejas contendo 15 kg de areia. Por ocasião da semeadura, as sementes foram submetidas a duas condições de disponibilidade hídrica (condição controle e condição estresse (deficit hídrico)). Para a condição controle, a areia foi umedecida com água até atingir a capacidade de campo. Para a condição estresse, uma solução de polietilenoglicol (PEG 6000) no potencial osmótico de -0,2 Mpa (Dantas et al., 2017) foi preparada conforme recomendação de Vilella et al. (1991), sendo esta solução posteriormente adicionada à areia.

Após o plantio, as bandejas foram devidamente pesadas, de modo que para cada bandeja foi anotado o peso de referência, para a umidade pretendida em cada condição hídrica. A irrigação foi realizada diariamente às 10h e 16h. A cada momento de irrigação, pesou-se todas as bandejas e aplicou-se o volume de água até que atingisse o peso de referência. O estresse por deficit hídrico foi imposto no momento da semeadura sendo finalizado aos 20 dias após o plantio. O potencial hídrico da folha foi verificado aos 15 dias após o plantio por meio de uma bomba de pressão (scholander) (Scholander et al., 1965).

A temperatura e umidade máxima e mínima foram acompanhadas diariamente. Na primeira época deste experimento, os valores médios para temperatura máxima e mínima foram de 32°C e 16 °C, respectivamente e umidade máxima e mínima de 95% e 42%, respectivamente. Já na segunda época, foram observados valores médios de 31°C e 15°C para temperatura máxima e mínima, respectivamente e 97% e 48% para umidade relativa máxima e mínima, respectivamente. Ressalta-se que cada bandeja alocou 10 parcelas experimentais, sendo que em cada parcela foram distribuídas 10 sementes. Para as variáveis inerentes à parte aérea e ao sistema radicular, quatro plantas centrais foram avaliadas, utilizando-se o valor médio destas.

Germinação de sementes

Avaliou-se a emergência das plântulas para obtenção do índice de velocidade de emergência (IVE). Considerou-se plântulas emergidas aquelas que apresentavam cotilédones acima da superfície do substrato. A contagem de plântulas emergidas foi realizada diariamente até o vigésimo dia de avaliação e o índice de velocidade de emergência foi calculado de acordo com Maguire (1962), sendo obtido pelo somatório das razões entre o número de plântulas verificadas no dia da contagem e o número de dias após a semeadura em que foi realizada a contagem. Aos 20 dias após o plantio, as plantas foram avaliadas quanto a emergência (%). Também foi avaliado o tempo para atingir 50% de emergência (T50), índice de uniformidade de emergência (IUG), taxa de emergência (TXMG), tempo médio de emergência (TMG), variância do tempo de emergência (VTG), coeficiente de variação do tempo de emergência (CVT), sincronia de emergência (SG), incerteza (IG) e velocidade do coeficiente de emergência (VCG). Todas estas últimas variáveis foram realizadas conforme descrito por Silva et al. (2019), sendo processados usando o software R e o pacote SeedCalc.

Conteúdo de água e massa seca

Passados os 20 dias após o plantio, de cada planta foram coletadas duas folhas unifolioladas. Este material foi devidamente embalado em papel alumínio, acondicionado em caixa de isopor resfriada e encaminhado ao laboratório para mensuração do conteúdo de água na folha. Isso foi feito subtraindo-se o peso da massa seca, do peso da massa fresca, dividindo o resultado pelo peso da massa fresca. Posteriormente, separou-se a parte aérea do sistema radicular das plantas, avaliando-se a altura das plântulas (cm) e massa seca da parte aérea (g). Esta última foi obtida após secagem do material vegetal em estufa de circulação forçada de ar a 65 °C, por 48 horas, procedendo-se com a posterior pesagem em balança de precisão.

Avaliação do sistema radicular

As raízes foram separadas da parte aérea, e após serem lavadas, foram armazenadas em álcool 70%, e acondicionadas em BOD (Biochemical Oxygen Demand) a 25°C. Posteriormente, as raízes foram avaliadas com o auxílio do software WinRHIZOPro 2009 (Basic, Reg, Pro and Arabidopsis for Root Measurement). Para tanto, cada raiz foi disposta em uma bandeja de 20 cm x 40 cm x 7 cm (largura x comprimento x altura) contendo água suficiente para cobrir as raízes, minimizando-se as sobreposições. As raízes foram digitalizadas com um scanner de mesa EPSON EU88 com resolução de 200 dpi. Por meio do processamento

das imagens obtidas foi possível obter as seguintes variáveis: comprimento total (cm), área superficial (cm²), diâmetro médio de raízes (mm), volume (cm³) e comprimento por classes de diâmetro (cm) (>0,5; >1,0; >1,5 mm). Posteriormente as raízes foram encaminhadas para secagem em estufa de circulação forçada de ar a 65°C, por 48 horas, procedendo-se com a posterior pesagem em balança de precisão, obtendo-se assim, a massa seca da raiz (g).

Experimento 2. Imposição de deficit hídrico no estágio fenológico v1

O experimento foi conduzido em casa de vegetação da Universidade Federal de Viçosa. Foram utilizadas sementes de 100 cultivares de soja (Tabela1), as quais foram tratadas com Cropstar (tiadicarbe e imidacloprido) + Maxim quatro (azoxistrobina, tiabendazol, fludioxonil e Metalaxil-M) e semeadas em vasos contendo 3 kg de substrato (solo + areia na proporção 2:1, respectivamente). Após a confecção do substrato, realizou-se a análise física e química para correção da fertilidade, a qual foi realizada de acordo com as exigências da cultura da soja.

Após o plantio, procedeu-se com irrigação de acordo com a demanda da cultura até que as plantas atingissem o estágio vegetativo v1 (folhas unifolioladas completamente desenvolvidas e a primeira folha trifoliolada com os bordos não mais se tocando (Fehr e Caviness (1977)). Quando as plantas atingiram este estágio fenológico, as mesmas foram submetidas a duas condições de disponibilidade hídrica (condição controle e condição estresse (deficit hídrico)). A condição controle consistiu em manter o solo em capacidade de campo (-10kPa). Já na condição estresse o solo foi mantido a tensão de 1200 kPa. A imposição do estresse por deficit hídrico foi feita utilizando-se a metodologia proposta por Nascimento et al. (2021). De forma resumida, essa metodologia requer a elaboração de uma curva de retenção de água no solo. A partir desta, calcula-se o conteúdo de água que o solo deve conter na condição controle (tensão de 10 kPa) e na condição estresse (tensão de 1200 kPa). O estresse por deficit hídrico foi mantido por 20 dias. Durante este período, uma amostra representativa de 20 vasos por bloco foi pesada diariamente às 10 e 16 horas, e por meio da diferença de massa do peso alvo do vaso e do peso do dia, o volume médio de água foi aplicado nos vasos a fim de atingir o potencial hídrico desejado. O potencial hídrico da folha foi verificado aos 15 dias após a imposição do deficit hídrico, por meio de uma bomba de pressão (scholander) (Scholander et al., 1965).

A temperatura máxima e mínima e umidade relativa do ar máxima e mínima foram acompanhadas diariamente durante o período de imposição do deficit hídrico para as duas épocas em que este experimento foi realizado. Na primeira época, os valores médios para

temperatura máxima foi 31° C e temperatura mínima foi 21°C, enquanto a umidade relativa máxima foi 87% e a mínima foi 31%. Já na segunda época, os valores foram os seguintes: 38° C e 19° C para temperatura máxima e mínima, respectivamente e 97% e 36% para umidade relativa.

Avaliações fenotípicas

Passados os 20 dias da imposição do estresse, procedeu-se com as avaliações das seguintes variáveis: altura (cm), número de nós, estágio vegetativo, diâmetro do hipocótilo (mm), massa seca da parte aérea (g), massa seca da raiz (g). Estas duas últimas variáveis foram obtidas conforme descrito no experimento 1. Também de forma semelhante ao realizado no experimento 1, avaliou-se as variáveis inerentes ao sistema radicular com auxílio do software Winrhizo. Foram elas: comprimento total (cm), área superficial (cm²), diâmetro médio de raízes (mm), volume (cm³), comprimento por classes de diâmetro (cm) (>0,5; >1,0; >1,5 mm).

Análises estatísticas

Os dados fenotípicos coletados nos dois experimentos das duas épocas foram analisados adotando-se a metodologia de modelos mistos via REML/BLUP (máxima verossimilhança restrita) (Patterson e Thompson, 1971), melhor predição linear não viesada (Henderson, 1975) a fim de estimar os componentes de variância, além de predizer os valores genotípicos.

Adotou-se o seguinte modelo estatístico:

$$y = X_m + Y_a + Z_c + W_b + T_i + U_j + \varepsilon$$

Em que:

y: é o vetor de dados; m é o vetor dos efeitos da combinação medição-repetição (assumida como fixa) adicionado à média geral; a é o vetor dos efeitos da época (assumido como fixo); g é o vetor de efeitos da cultivar (assumido como aleatório); b é o vetor de efeitos ambientais (assumidos como aleatórios); i é o vetor dos efeitos cultivar × ambiente (assumido como aleatório); j é o vetor dos efeitos cultivar × época (assumido como aleatório); ε é o vetor do resíduo (aleatório); e X, Y, Z, W, T e U representam as matrizes de incidência para esses efeitos.

Para os efeitos aleatórios do modelo, a significância do teste de razão de verossimilhança foi avaliada por meio da estatística qui-quadrado com um grau de liberdade. Os valores genéticos (médias BLUP) foram preditos para cada uma das 100 cultivares de soja

considerando-se as variáveis que apresentaram efeito significativo para o fator cultivar, avaliadas nos dois experimentos, nas duas épocas em que foram realizadas.

Índice de seleção FAI-BLUP

Para ranquear as cultivares quanto a tolerância ao déficit hídrico foi utilizado o índice de seleção FAI-BLUP (Rocha et al., 2018). Esse índice utiliza os valores genéticos (médias BLUP), baseando-se na análise fatorial e no desenho ideótipo-genótipo, para realizar o ranqueamento. A análise de componentes principais foi utilizada para extrair as cargas fatoriais da matriz de correlação genética obtida a partir dos valores genéticos. Para a rotação analítica foi utilizado o critério varimax (Kaiser, 1958), enquanto para o cálculo dos escores fatoriais foi utilizado o método dos mínimos quadrados ponderados, conforme proposto por Bartlett (1978). Adotou-se a intensidade de seleção de 15%.

Os ideótipos de soja para tolerância ao déficit hídrico foram desenhados considerando os valores máximos ou mínimos para as características avaliadas (Tabela 2).

Tabela 2. Valores máximos e mínimos para as variáveis utilizados no desenho do ideótipo.

Variável	Máx./Mín	Variável	Máx./Mín.
Emergência	Máximo	Diâmetro do hipocótilo	Máximo
Índice de vel. de emergência	Máximo	Número de nós	Máximo
Tempo para 50% de emergência	Mínimo	Estádio fenológico	Máximo
Índice de unif. de emergência	Máximo	Peso seco da parte aérea	Máximo
Tempo médio de emergência	Mínimo	Peso seco de raiz	Máximo
Taxa média de emergência	Máximo	Comprimento de raiz	Máximo
Var. do tempo de emergência	Mínimo	Area projetada	Máximo
CV do tempo de emergência	Mínimo	Area superficial	Máximo
Sincronia de emergência	Máximo	Diâmetro médio de raiz	Mínimo
Incerteza de emergência	Mínimo	Volume de raiz	Máximo
Vel. do coeficiente de emergência	Máximo	CD1	Mínimo
Conteúdo de água	Máximo	CD2	Mínimo
Altura	Máximo	CD3	Mínimo

Var. variância; CV= coeficiente de variação; CD1 = comprimento de raiz por classe de diâmetro (>0,5); CD2 = comprimento de raiz por classe de diâmetro (>0,5<1,0); CD3 = comprimento de raiz por classe de diâmetro (>1,0 <1,5 mm).

Para definição do número de ideótipos considerou-se uma combinação de fatores desejáveis e indesejáveis, utilizando-se o seguinte algoritmo:

$$NI=2^n.$$

Em que: NI é o número de ideótipos e n é o número de fatores.

Posteriormente a determinação do ideótipo, foram estimadas as distâncias genótipo-ideótipo as quais foram convertidas em probabilidade espacial permitindo assim, o ranqueamento das cultivares em estudo. Para tanto, o seguinte algoritmo foi usado:

$$P_{ij} = \frac{\frac{1}{d_{ij}}}{\sum_{i=1; j=1}^{i=n; j=m} \frac{1}{d_{ij}}}$$

Em que:

P_{ij} = Probabilidade do i -ésimo genótipo ($i = 1, 2, \dots, n$) ser semelhante ao j -ésimo ideótipo ($j = 1, 2, \dots, m$); d_{ij} = Distância genótipo-ideótipo do i -ésimo genótipo ao j -ésimo ideótipo com base na distância euclidiana média padronizada.

Índice de seleção MGIDI

O ranqueamento de cultivares de soja quanto à tolerância ao deficit hídrico utilizando o índice de seleção MGIDI se deu conforme recomendações de Olivoto e Nardino (2021), da seguinte forma:

$$MGIDI_i = [\sum_{j=1}^f (y_{ij} - y_j)^2]^{0.5}$$

Em que:

$MGIDI_i$ = é o índice multicaestirística de distância genótipo-ideótipo; y_{ij} = é a pontuação do i -ésimo genótipo ($i=1,2,3\dots g$) no j -ésimo fator ($f=1,2,3\dots f$), sendo g e f o número de genótipos e fatores, respectivamente; e y_j é a j -ésima pontuação do ideótipo. O genótipo com menor MGIDI fica então mais próximo do ideótipo e, portanto, apresenta valores desejados para todas as características.

Também foi realizada a proporção do índice MGIDI do genótipo explicada pelo fator j (ω_{ij}), o qual foi usado para mostrar os pontos fortes e fracos dos genótipos sendo calculada como:

$$\omega_{ij} = \frac{\sqrt{D_{ij}^2}}{\sum_{j=1}^f \sqrt{D_{ij}^2}}$$

Em que:

D_{ij} = é a distância entre o i -ésimo genótipo e o ideótipo para o j -ésimo fator. Nesse caso, baixas contribuições de um fator indicam que as variáveis dentro de tal fator estão próximas do ideótipo.

Foi adotada a intensidade de seleção de 15%. As análises foram realizadas no software R, utilizando os pacotes Metan, readx1 e sommer. Abaixo, encontra-se um resumo da metodologia utilizada neste estudo (Figura 1).

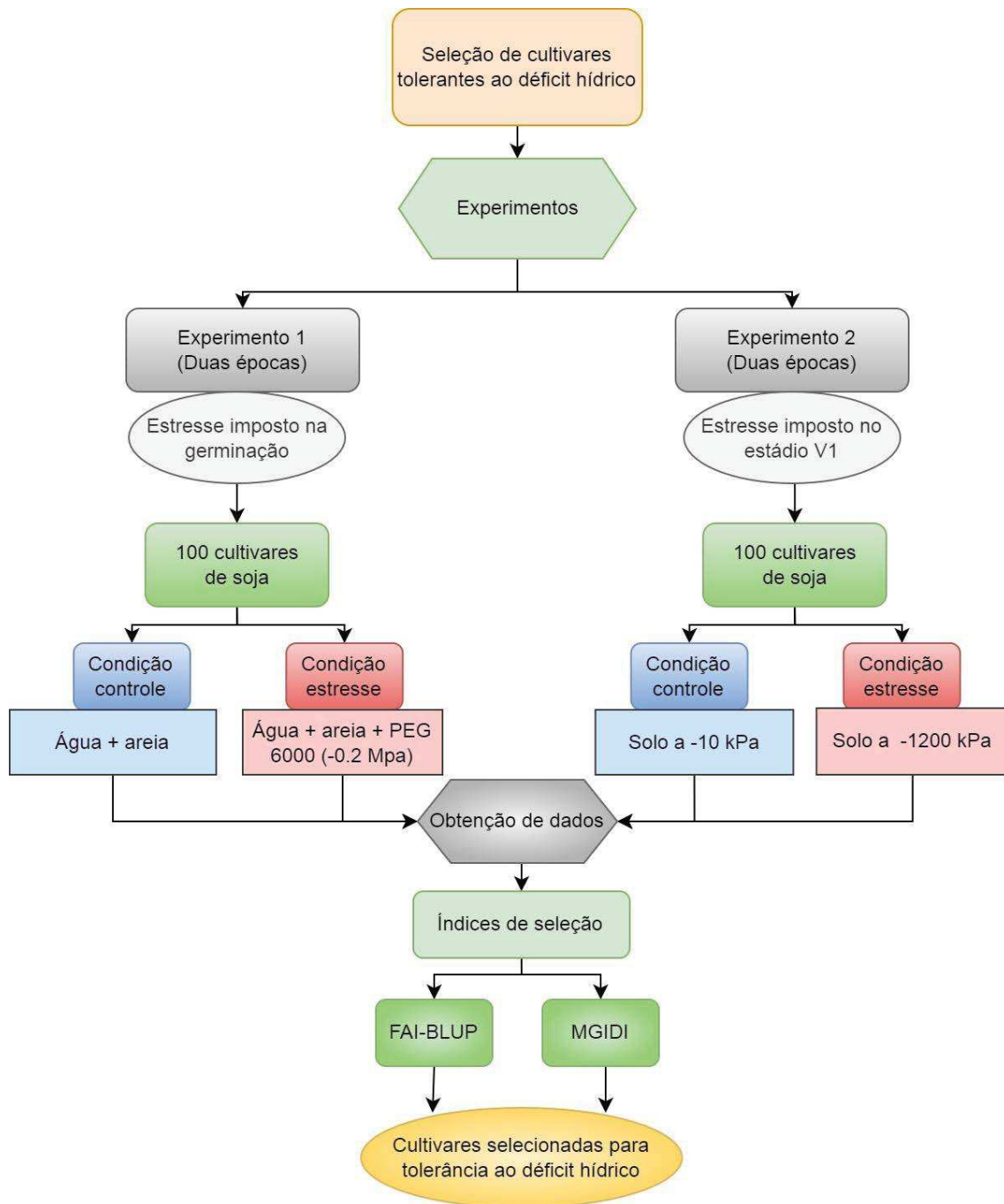


Figura 1. Resumo da metodologia utilizada no estudo.

3. RESULTADOS E DISCUSSÃO

Nem todas as variáveis avaliadas apresentaram significância para o fator cultivar. No experimento 1, no qual o estresse por déficit hídrico foi imposto na fase de germinação, apenas a variável altura apresentou efeito significativo nas duas épocas do experimento. Já no experimento 2, em que o déficit hídrico foi imposto no estágio fenológico V1, para ambas as épocas, as seguintes variáveis apresentaram efeito significativo para o fator cultivar: altura, diâmetro do hipocótilo, número de nós, peso seco de raiz, estágio fenológico, área projetada da raiz, área superficial da raiz, comprimento por volume de raiz, comprimento de raiz com diâmetro menor que 0,5, comprimento de raiz com diâmetro maior que 0,5 e menor que 1,0 e comprimento de raiz com diâmetro maior que 1,0 e menor que 1,5 mm. Além destas, na segunda época deste experimento, as variáveis diâmetro do hipocótilo e peso seco da raiz também apresentaram efeito significativo para o fator cultivar.

Os autovalores e a variância acumulada dos componentes principais obtidos pela matriz de correlação genética estão apresentados na Tabela 3. Observa-se que os quatro primeiros componentes apresentaram autovalores superiores a um (7,33; 2,15; 1,51; 1,13, para os componentes 1, 2, 3 e 4, respectivamente). Esse resultado sugere a possibilidade de redução de dimensionalidade dos dados para apenas quatro fatores, seguindo o critério de Kaiser (1958). De acordo com este critério, o número de fatores é definido pelo número de autovalores iguais ou superiores à unidade. A variância acumulada dos primeiros quatro componentes principais foi responsável por aproximadamente 80.91% de toda a variabilidade genética presente no conjunto de dados.

Tabela 3. Estimativas dos autovalores da análise de componentes principais e a proporção da variância explicada por cada um deles.

Componente principal (PC)	Autovalores	Variância acumulada
PC1	7.33	48.86
PC2	2.15	63.20
PC3	1.51	73.32
PC4	1.13	80.91
PC5	0.95	87.29
PC6	0.66	91.72
PC7	0.49	94.99
PC8	0.32	97.16
PC9	0.24	98.76
PC10	0.10	99.47
PC11	0.51	99.81
PC12	0.02	99.98
PC13	0.002	99.99
PC14	6.79E-05	100
PC15	9.39E-14	100

As variáveis utilizadas neste estudo foram agrupadas em quatro fatores (Figura 2). Cada fator é composto por um conjunto de variáveis que apresenta elevada correlação entre si e baixa correlação com as demais variáveis que compõem outros fatores (Cruz e Carneiro, 2003). O fator 1 incluiu basicamente as variáveis radiculares, sendo elas peso seco de raiz, área projetada, área superficial, comprimento por volume de raiz, volume de raiz, comprimento de raiz com diâmetro menor que 0,5 mm, comprimento de raiz com diâmetro maior que 0,5 e menor que 1,0 mm e comprimento de raiz com diâmetro maior que 1,0 mm e menor que 1,5 mm. No fator 2, foram incluídas as variáveis diâmetro do hipocótilo e peso seco de raiz, sendo ambas da segunda época do experimento 2. O fator 3, agrupou as variáveis altura, número de nós e estágio fenológico. Já o fator 4, alocou a variável altura do experimento 1 e diâmetro do hipocótilo do experimento 2, ambas da primeira época dos experimentos. Ressalta-se que na análise fatorial, variáveis que são agrupadas em um fator têm maior correlação entre si do que variáveis de diferentes fatores.

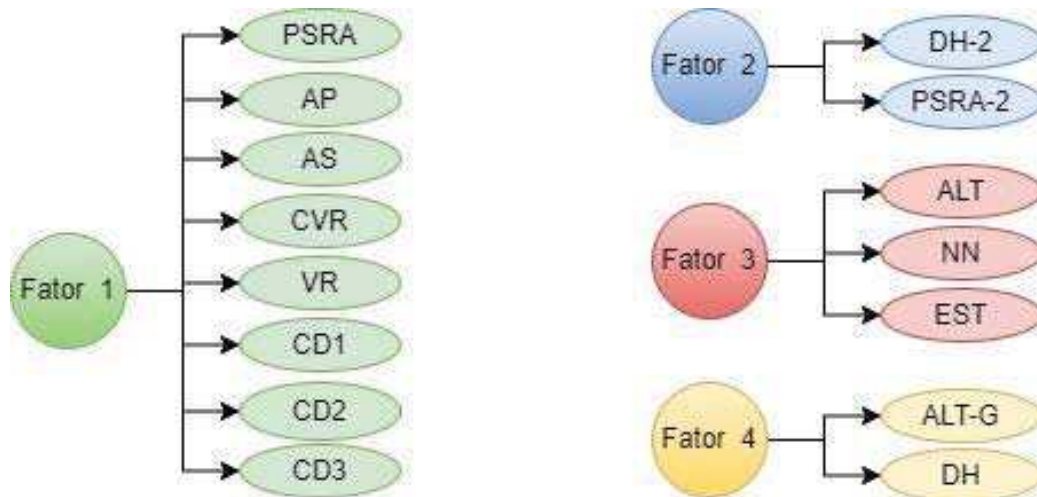


Figura 2. Agrupamento das variáveis em fatores.

PSRA = peso seco de raiz; AP = área projetada; AS = área superficial, CVR = comprimento por volume de raiz, VR = volume de raiz, CD1 = comprimento de raiz por classe de diâmetro ($>0,5$); CD2 = comprimento de raiz por classe de diâmetro ($>0,5 < 1,0$); CD3 = comprimento de raiz por classe de diâmetro ($>1,0 < 1,5$); DH-2 = diâmetro do hipocótilo do experimento 2 na segunda época; PSRA-2 = peso seco de raiz do experimento 2 na segunda repetição; ALT = altura; NN = número de nós; EST = estágio fenológico; ALT-G = altura do experimento 1; DH = diâmetro do hipocótilo do experimento 2 na primeira época. Todas as variáveis incluídas no fator 1 e 3 foram obtidas do experimento 2 na primeira época.

As cargas fatoriais após a rotação varimax para os quatro fatores encontram-se na Figura 3. As cores indicam a correlação entre as variáveis no fator. Dessa forma, quanto mais acentuada a tonalidade do azul, indica ocorrência de correlação mais negativa. De forma semelhante, quanto mais vermelha, indica ocorrência de correlação mais positiva. Observou-se elevadas correlações entre as variáveis dentro de todos os quatro fatores. As variáveis que apresentaram colorações mais intensas foram as que apresentaram maiores correlações entre si dentro de um mesmo fator. Por exemplo, no fator 3, as variáveis altura e número de nós apresentam maior correlação entre si e menor correlação com a variável estágio fenológico.

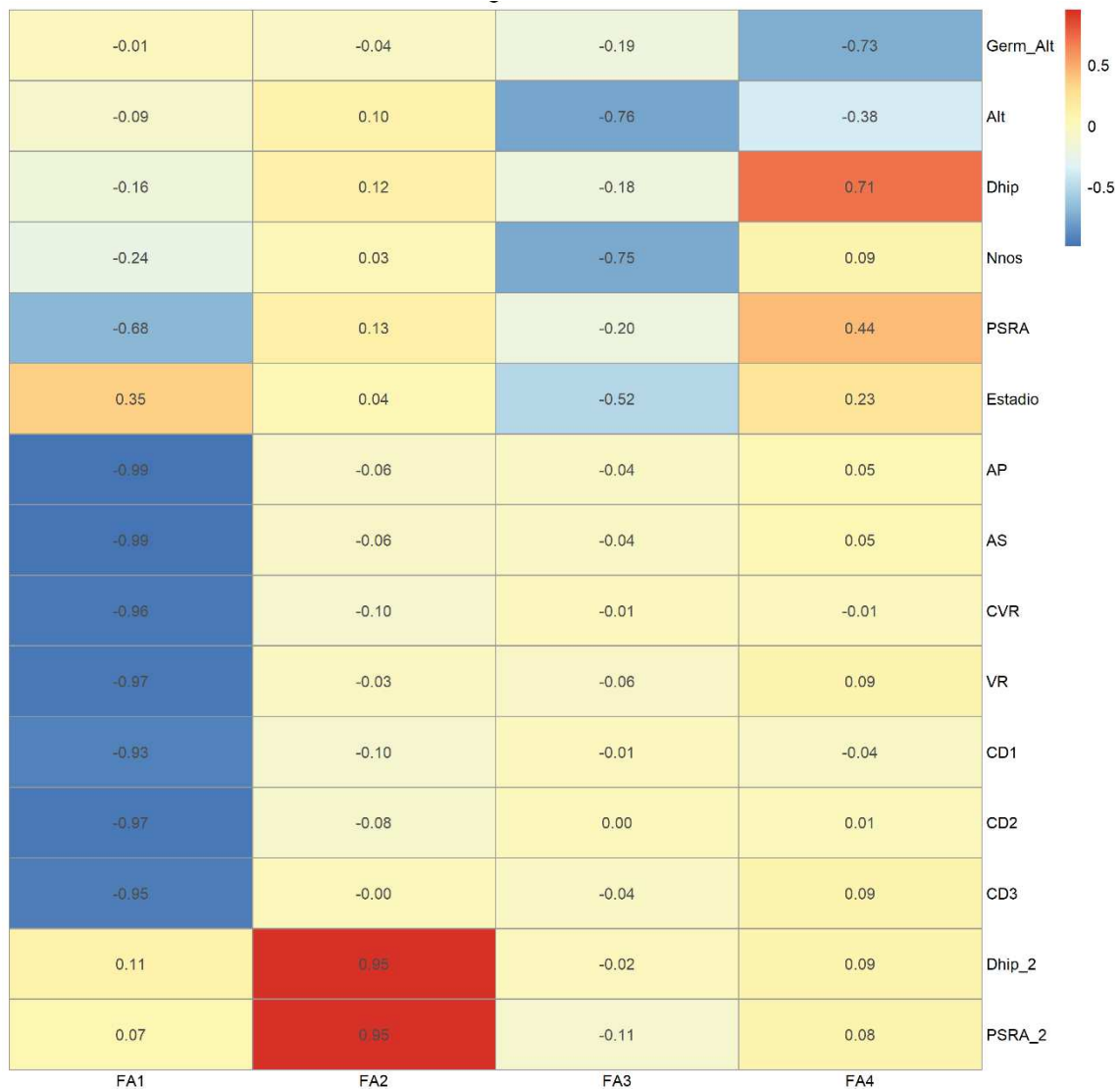


Figura 3. Mapa de calor mostrando as cargas fatoriais após a rotação varimax para os fatores. PSRA = peso seco de raiz; AP = área projetada; AS = área superficial, CVR = comprimento por volume de raiz, VR = volume de raiz, CD1 = comprimento de raiz por classe de diâmetro (>0,5); CD2 = comprimento de raiz por classe de diâmetro (>0,5<1,0); CD3 = comprimento de raiz por classe de diâmetro (>1,0 <1,5)); Dhip-2 = diâmetro do hipocótilo do experimento 2 na segunda época; PSRA-2 = peso seco de raiz no experimento 2 na segunda repetição; ALT = altura; NN = número de nós; Germ-alt = altura da planta no experimento 1; DH = diâmetro do hipocótilo no experimento 2 na primeira época. Todas as variáveis incluídas no fator 1 e 3 foram obtidas do experimento 2 na primeira época.

Ranqueamento de cultivares utilizando o índice de seleção FAI-BLUP

Um ideótipo é definido a partir de um conjunto de características agronômicas desejáveis (Woyann et al., 2019). Por isso, o ideótipo de cultivares de soja visando tolerância

ao deficit hídrico foi desenhado levando em consideração as características desejáveis para os quatro fatores. O ranqueamento das cultivares utilizando o índice de seleção FAI-BLUP, bem como a probabilidade das cultivares apresentarem tolerância ao deficit hídrico em relação a distância do ideótipo, pode ser observado na Figura 4. As cultivares selecionadas foram as seguintes M 9144RR, 5D 660 RR, NK 7059 RR, BMX TITAN RR, CD 243 RR, CD 2792 RR, OCEPAR 3 PRIMAVERA, CD 202, BRAGG, CD 236 RR, P98Y70, CD 240 RR, FUNDACEP 53 RR, A 6001 RR, EMGOPA 304 (Campeira). Ressalta-se que as cultivares M 9144 RR e 5D 660 RR foram as que mais se aproximaram do ideótipo desejável. Isso sugere que essas cultivares sejam priorizadas nos programas de melhoramento de soja visando a tolerância ao deficit hídrico, já que apresentam fenótipos favoráveis para as características de interesse.

As cultivares BMX TITAN RR e P 98Y70 também foram selecionadas como tolerantes ao deficit hídrico em um estudo (Nascimento, 2020), no qual o estresse por deficit hídrico foi imposto na fase reprodutiva (estádio R5) permanecendo por 15 dias. Dessa forma, a permanência dessas cultivares em estudos futuros é fortemente encorajada, uma vez que ao utilizar o índice FAI-BLUP, foram selecionadas como tolerantes ao deficit hídrico no estágio inicial de desenvolvimento e também na fase reprodutiva. Estes resultados sugerem que ambas as cultivares, de fato, apresentam grande potencial de tolerância ao deficit hídrico.

As cultivares P98Y11, BRSMT (PINTATO), A 8000, NA 4990 RG, FUNDACEP 56 RR, CD 234 RR, CD 250 RR, SYN 3358 RR, CD 5807, CD 246, ANTA 82, FUNDACEP 58 RR, M-SOY 6101, BRS 245 RR, CD 254 RR, apresentaram-se mais distantes do ideótipo para as características avaliadas. Essas cultivares, por não apresentarem proximidade com o ideótipo, não devem ser incluídas nos estudos futuros visando a tolerância ao deficit hídrico.

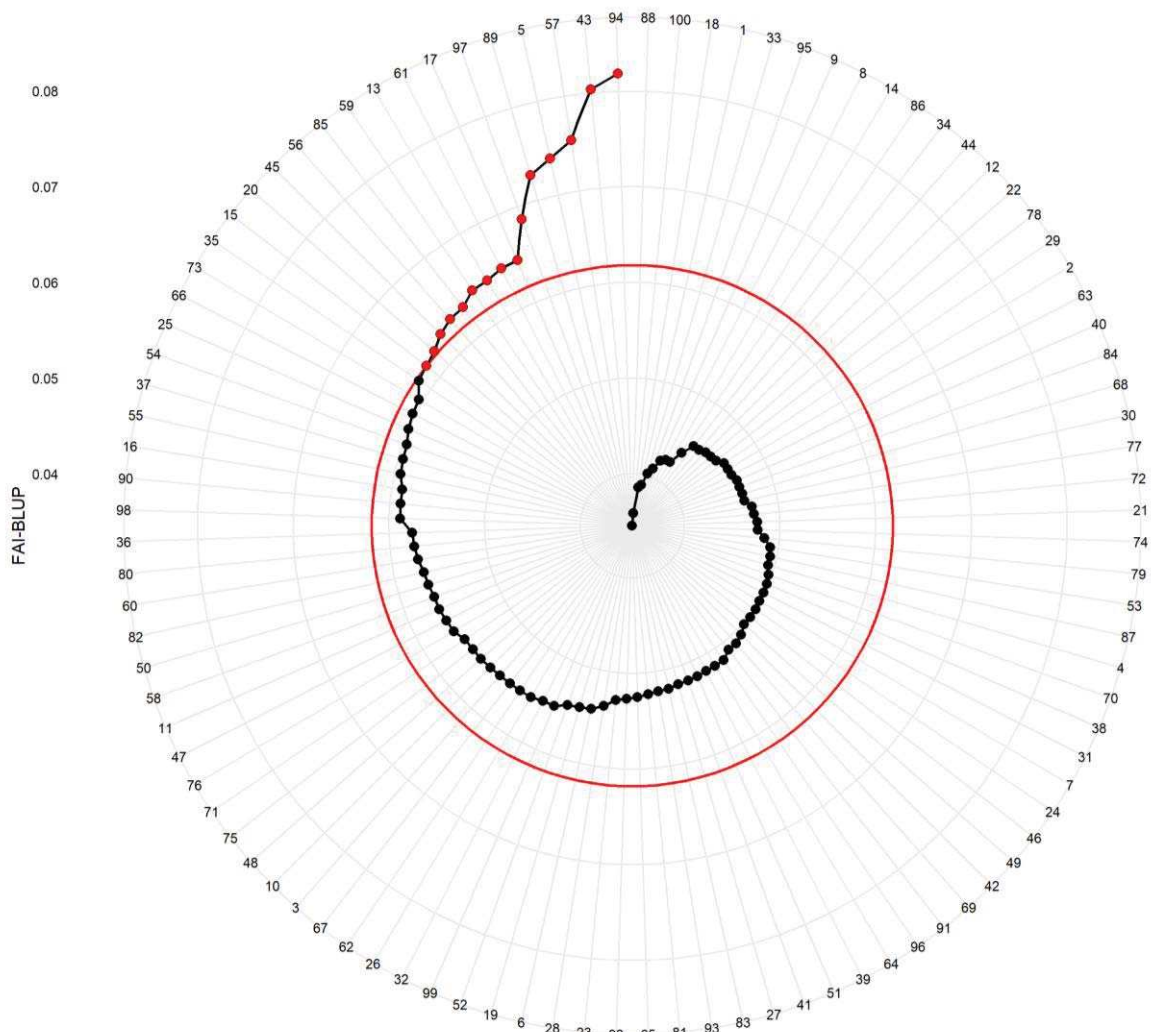


Figura 4. Ranqueamento de cultivares de soja quanto a tolerância ao deficit hídrico utilizando o índice FAI-BLUP. Os pontos de cor vermelha indicam as cultivares selecionadas. O círculo vermelho representa o ponto de corte de acordo com a intensidade de seleção.

O índice de seleção FAI-BLUP possibilitou ranquear as cultivares de soja quanto à tolerância ao deficit hídrico, considerando as múltiplas características e as duas condições de disponibilidade hídrica simultaneamente. Dessa forma, a seleção baseada no ideótipo, considerando as informações de cada ambiente, potencializa os efeitos das interações genótipos \times ambiente. Ademais, esse índice possui a vantagem de ser livre de problemas decorrentes de multicolinearidade (Rocha et al., 2018).

Woyann et al. (2019) utilizaram o índice FAI-BLUP para selecionar linhagens de soja para produção de biodiesel. Os autores conseguiram selecionar seis linhagens recombinantes adequadas para a produção de biodiesel, as quais apresentaram um conjunto de características mais próximo do ideótipo para a característica de interesse. Os autores concluíram que este

índice de seleção é eficiente para seleção e melhoramento de soja, contribuindo com incrementos genéticos significativos nos programas de melhoramento. Rocha et al. (2019) utilizaram o índice FAI-BLUP para selecionar genótipos de feijão, e, concluíram que as progênes indicadas pelo índice FAI-BLUP contribuíram para acelerar o desenvolvimento de genótipos superiores.

Ranqueamento de cultivares de soja utilizando o índice de seleção MGIDI

Das quinze cultivares selecionadas pelo índice MGIDI, apenas três não são comuns àquelas selecionadas pelo índice FAI-BLUP. As cultivares selecionadas pelo índice MGIDI (Figura 5) ocuparam a seguinte posição: M9144RR, 5D 660 RR, BMX TITAN RR, CD 2792 RR, NK 7059 RR, CD 240 RR, EMGOPA 304 (Campeira), OCEPAR 3 PRIMAVERA, CD 243 RR, CD 233 RR, CD 202, FUNDACEP 53 RR, 5D 711 RR, BRS 232 e P98Y70. Apenas as cultivares CD 233 RR, 5D 711 RR e BRS 232 diferiram daquelas selecionadas pelo índice anterior.

As cultivares M9144RR, 5D 660 RR ocuparam a mesma posição no ranqueamento realizado pelos dois índices utilizados, apresentando maior proximidade com o ideótipo desejado. Esse resultado observado em ambos os índices de seleção utilizados neste estudo reforça a importância de priorizar estas duas cultivares em estudos futuros nos programas de melhoramento de soja visando tolerância ao déficit hídrico. Em relação as cultivares que mais se distanciaram do ideótipo, o índice MGIDI classificou as seguintes cultivares: P98Y11, A 8000, BRSMT (PINTATO), FUNDACEP 56 RR, NA 4990 RG, ANTA 82, CD 217, CD 5807, CD 250 RR, M-SOY 6101, SYN 3358 RR, CD 235 RR, CD 246 e CD 234RR.

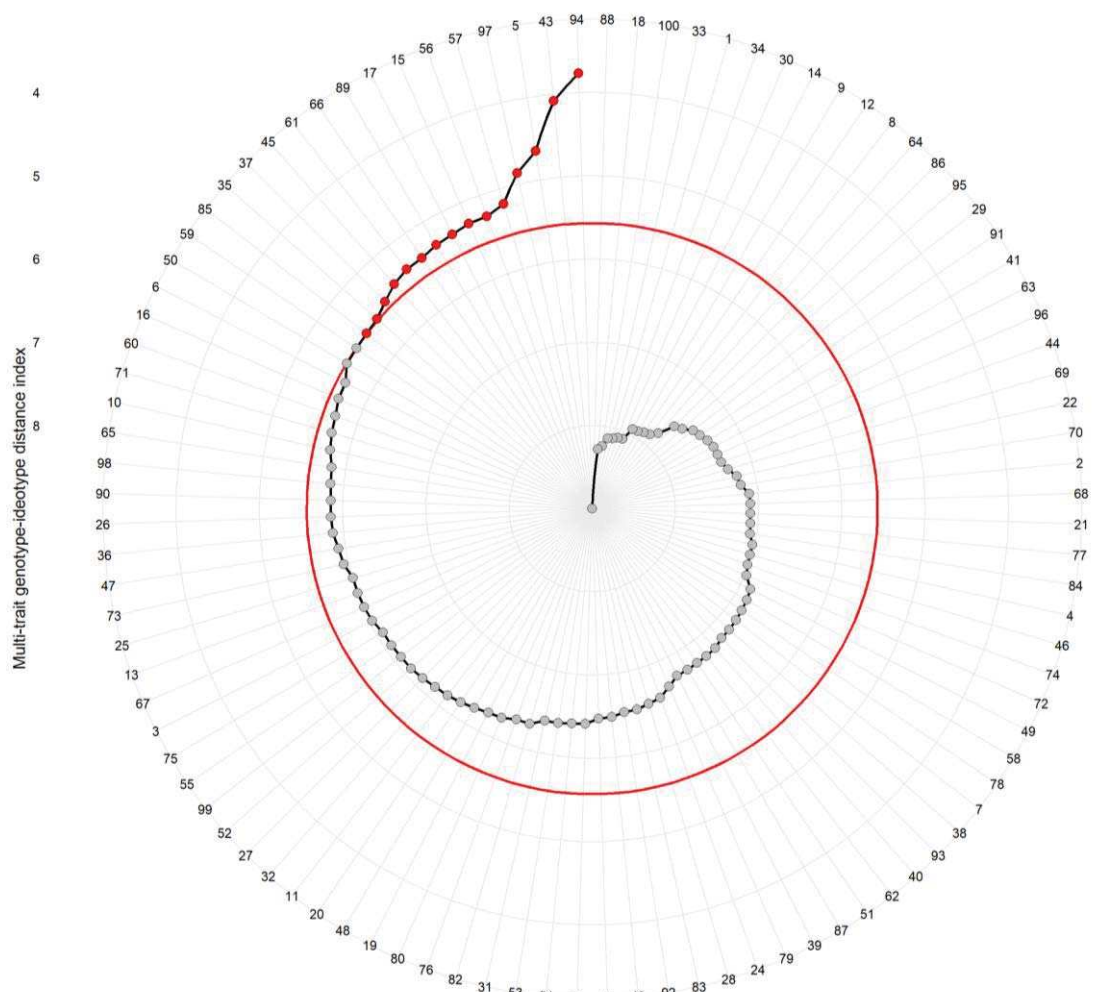


Figura 5. Ranqueamento de cultivares de soja quanto a tolerância ao déficit hídrico utilizando o índice MGIDI. Os pontos de cor vermelha indicam as cultivares selecionadas. O círculo vermelho representa o ponto de corte de acordo com a intensidade de seleção.

Os pontos fortes e fracos das cultivares, que são contabilizados pela proporção de cada fator para o índice MGIDI são apresentados na Figura 6. Quanto menor a proporção explicada por um fator (mais próximo da borda externa), mais próximos as variáveis desse fator estão do ideótipo (Olivoto e Nardino, 2021). Portanto, o FA3 apresentou a maior contribuição, ao passo que o FA4 apresentou a menor contribuição para o índice MGIDI. A linha tracejada mostra o valor teórico em que todos os fatores contribuíram igualmente.

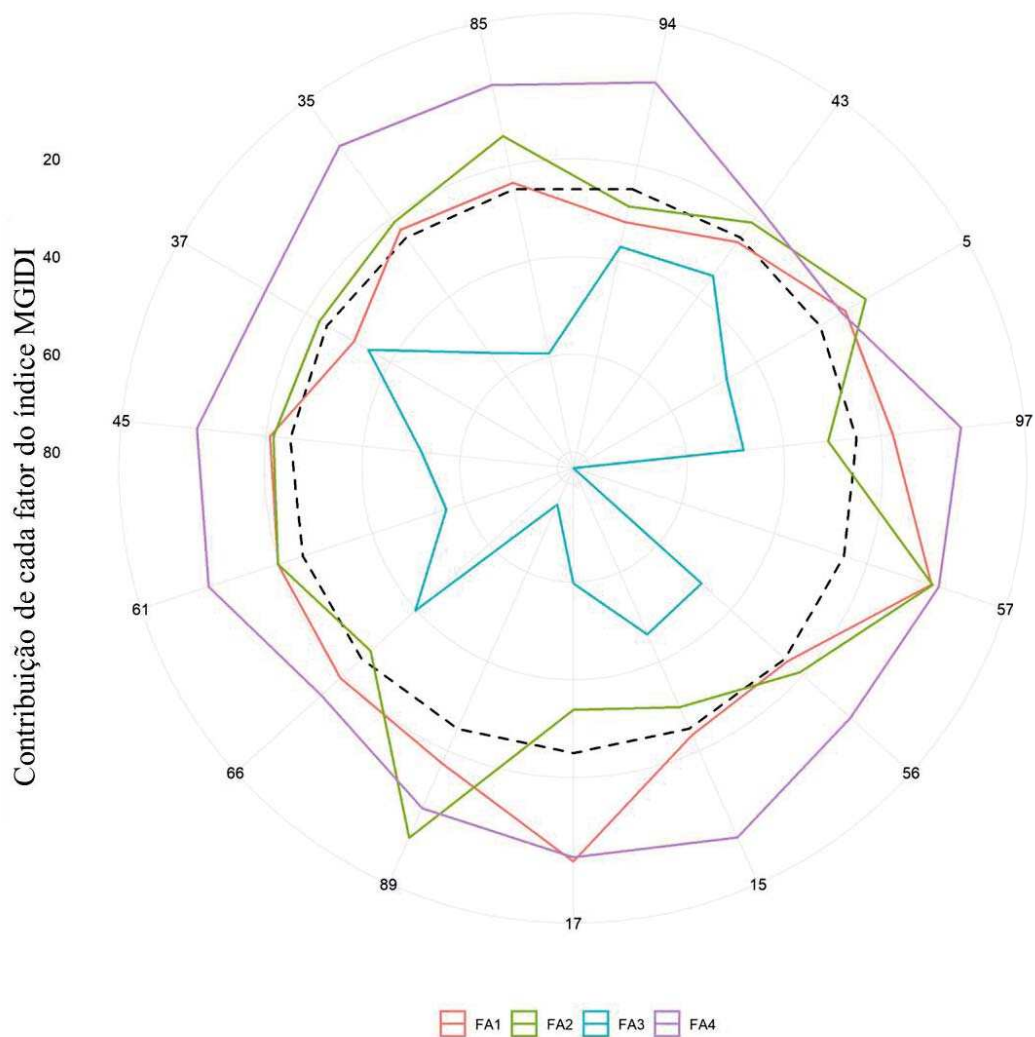


Figura 6. Pontos fortes e fracos das cultivares selecionadas pelo índice MGIDI contabilizados pela proporção de cada fator (FA).

Observa-se que o FA3 apresentou a menor contribuição para as cultivares 5D 711 RR(37), M 9144 RR(94), e 5D 660 RR (43). Esse resultado indica que dentre as cultivares selecionadas, essas foram as mais desenvolvidas, considerando-se os valores das variáveis altura, número de nós e estágio fenológico. Por outro lado, o FA3 apresentou a maior contribuição do MGIDI da cultivar NK 7059 RR (57), sugerindo que esta cultivar apresentou ou menores valores para as variáveis supramencionadas.

O FA1 incluiu as variáveis inerentes a morfologia de raiz. Esse fator apresentou menores contribuições para as cultivares OCEPAR 3 PRIMAVERA (17) e NK 7059 RR (57).

Possivelmente essas cultivares apresentaram sistema radicular mais avantajado em relação as demais cultivares selecionadas. O maior desenvolvimento do sistema do radicular, com raízes mais profundas, ocupando maior área no solo, bem como com ocorrência de raízes mais finas é uma boa estratégia para lidar com restrição hídrica. Os resultados deste estudo indicaram que a cultivar NK 7059 RR (57) apresentou menor desenvolvimento da parte aérea. Entretanto, esta mesma cultivar parece ter compensado essa deficiência com maior desenvolvimento de raízes. Isso pode ter garantido que esta cultivar ocupe posição próxima do ideótipo no ranqueamento realizado pelos dois índices de seleção.

Raízes finas, conseguem explorar maior volume de solo, e por isso conseguem absorver maior quantidade de água e nutrientes (Smith e Read 2008), bem como promover um maior crescimento de raízes (Ola et al. 2015). De acordo com Bardgett et al. (2014) e Poorter et al. (2012) a plasticidade da raiz é essencial para superar e lidar com o estresse por déficit hídrico. Ademais, as plantas que conseguem alterar a estrutura das raízes, como estratégia para lidar com a seca, podem explorar a água em camadas mais profundas do solo, permitindo uma melhoria acentuada no rendimento de grãos (Uga et al. (2013).

Dado o exposto, ressalta-se que os resultados encontrados no presente estudo muito podem contribuir com pesquisas futuras dos programas de melhoramento de soja visando a tolerância ao deficit hídrico. A seleção de cultivares que apresentem potencial de tolerância à esse tipo de estresse abiótico na fase inicial de desenvolvimento é essencial para avanços na busca por materiais tolerantes, já que se trata de uma fase importante para o desenvolvimento e estabelecimento da lavoura (Hua et al., 2018; Zhang et al., 2022). Ademais, as cultivares de soja selecionadas podem ser utilizadas como fonte de germoplasma em estudos futuros nos programas de melhoramento visando a tolerância ao deficit hídrico. Por fim, a realização de estudos nessa fase apresenta como principais vantagens o fato de apresentar baixo custo, além de ser de mais fácil execução e requerer um ciclo experimental mais curto.

4. CONCLUSÕES

Das 15 cultivares selecionadas pelos índices FAI-BLUP e MGIDI, 12 delas foram comuns entre os índices de seleção.

As cultivares M 9144 RR, BMX TITAN RR foram as que mais se aproximaram do ideótipo em ambos os índices de seleção.

5. REFERÊNCIAS

- Ahanger, M. A., Siddique, K. H., & Ahmad, P. (2021). Understanding drought tolerance in plants. *Physiol Plant*, *172*, 286-8.
- Ahmed, I., Rohman, M., Hossain, M., Molla, M., Azam, M., Hasan, M., ... & Hossain, A. (2022). A Study on the Phenotypic Variation of 103 Cucumber (*Cucumis sativus* L.) Landraces for the Development of Desirable Cultivars Suitable for the Changing Climate. *Life*, *12*(8), 1235.
- Bermudez, F., & Pinheiro, J. B. (2020). Selection to high productivity and stink bugs resistance by multivariate data analyses in soybean. *Bragantia*, *79*, 250-259.
- Bewley, J. D., & Black, M. (2012). *Physiology and biochemistry of seeds in relation to germination: volume 2: viability, dormancy, and environmental control*. Springer Science & Business Media.
- CRUZ, C.D., CARNEIRO, P.C.S. Modelos Biométricos aplicados ao melhoramento genético. Viçosa: UFV, v.2, 2003.
- da Silva Junior, A. C., de Castro Sant'Anna, I., Peixoto, M. A., Torres, L. G., Silva Siqueira, M. J., da Costa, W. G., ... & Cruz, C. D. (2022). Multiple-trait model through Bayesian inference applied to flood-irrigated rice (*Oryza sativa* L). *Euphytica*, *218*(9), 1-12.
- Dantas, S. A. G., Silva, F. C. S., Silva, L. J., & Silva, F. L. (2017). Strategy for selection of soybean genotypes tolerant to drought during germination. *Genetics and Molecular Research* *16* (2).
- Du, Y., Zhao, Q., Chen, L., Yao, X., & Xie, F. (2020). Effect of drought stress at reproductive stages on growth and nitrogen metabolism in soybean. *Agronomy*, *10*(2), 302.
- Fathi, A., & Tari, D. B. (2016). Effect of drought stress and its mechanism in plants. *International Journal of Life Sciences*, *10*(1), 1-6.
- Fehr, W., & Caviness, C. (1977). Stages of soybean development.

- Hameed, A., El-Keblawy, A., Aljasmi, M., Gairola, S., Phartyal, S. S., Mosa, K. A., & Soliman, S. (2021). Seed provenance, thermoperiod, and photoperiod affect low water potential tolerance during seed germination of the multipurpose exotic tree *Prosopis juliflora*. *Journal of Arid Environments*, *195*, 104627.
- Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics*, *28*(6), 476-490.
- He, J., Du, Y. L., Wang, T., Turner, N. C., Yang, R. P., Jin, Y., ... & Li, F. M. (2017). Conserved water use improves the yield performance of soybean (*Glycine max* (L.) Merr.) under drought. *Agricultural Water Management*, *179*, 236-245.
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(3), 187-200.
- Khatun, M., Sarkar, S., Era, F. M., Islam, A. M., Anwar, M. P., Fahad, S., ... & Islam, A. A. (2021). Drought stress in grain legumes: Effects, tolerance mechanisms and management. *Agronomy*, *11*(12), 2374.
- Kistner, M. B., Nazar, L., Montenegro, L. D., Cervigni, G. D. L., Galdeano, E., & Iglesias, J. (2022). Detecting sources of resistance to multiple diseases in Argentine maize (*Zea mays* L.) germplasm. *Euphytica*, *218*(5), 1-18.
- Leite, W. D. S., Unêda-Trevisoli, S. H., Silva, F. M. D., Silva, A. J. D., & Mauro, A. O. D. (2018). Identification of superior genotypes and soybean traits by multivariate analysis and selection index. *Revista Ciência Agronômica*, *49*, 491-500.
- Leite, W. D. S., Unêda-Trevisoli, S. H., Silva, F. M. D., Silva, A. J. D., & Mauro, A. O. D. (2018). Identification of superior genotypes and soybean traits by multivariate analysis and selection index. *Revista Ciência Agronômica*, *49*, 491-500.
- Marinho, J. P., Pagliarini, R. F., Molinari, M. D. C., Marcolino-Gomes, J., Caranhoto, A. L. H., Marin, S. R. R., ... & Mertz-Henning, L. M. (2022). Overexpression of full-length and partial DREB2A enhances soybean drought tolerance. *Agronomy Science and Biotechnology*, *8*, 1-21.
- MARTINS, D. (2019). Biochemical examination of non-transgenic and transgenic soybean plants under drought stress conditions. *Biologia plantarum*, *63*, 314-322.
- Mengistu, A., Ray, J. D., Smith, J. R., Arelli, P. R., Bellaloui, N., Chen, P., ... & Boykin, D. (2018). Effect of charcoal rot on selected putative drought tolerant soybean genotypes and yield. *Crop Protection*, *105*, 90-101.

- Nascimento, H. R. D. (2020). Identificação de caracteres e seleção de genótipos de soja tolerantes ao deficit hídrico.
- Nascimento, H. R., Moura, L. D. O., Duarte, A. B., Dantas, S. A. G., Ferreira, D. D. O., Rosmaninho, L. B. D. C., ... & da Silva, F. L. (2021). A new methodological approach for simulating water deficit in soybean genotypes. *Journal of Agronomy and Crop Science*, 207(6), 946-955.
- Olivoto, T., & Nardino, M. (2021). MGIDI: toward an effective multivariate selection in biological experiments. *Bioinformatics*, 37(10), 1383-1389.
- Olivoto, T., & Nardino, M. (2021). MGIDI: toward an effective multivariate selection in biological experiments. *Bioinformatics*, 37(10), 1383-1389.
- Pathy, T. L., & Mohanraj, K. (2021). Estimating best linear unbiased predictions (BLUP) for yield and quality traits in sugarcane. *Sugar Tech*, 23(6), 1295-1305.
- Pessoa, H. P., Da Silva, L. J., Chagas, R. R., Copati, M. G. F., Dariva, F. D., de Oliveira Dias, F., & Gomes, C. N. (2021). Uncovering tomato candidate genes for seed germination performance and seedling vigour using the *Solanum pennellii* introgression line population. *Scientia Horticulturae*, 284, 110134.
- Pour-Aboughadareh, A., Sanjani, S., Nikkiah-Chamanabad, H., Mehrvar, M. R., Asadi, A., & Amini, A. (2021). Identification of salt-tolerant barley genotypes using multiple-traits index and yield performance at the early growth and maturity stages. *Bulletin of the National Research Centre*, 45(1), 1-16.
- Rangel, R. M., Amaral Júnior, A. T. D., Gonçalves, L. S. A., Freitas Júnior, S. D. P., & Candido, L. S. (2011). Análise biométrica de ganhos por seleção em população de milho pipoca de quinto ciclo de seleção recorrente. *Revista Ciência Agronômica*, 42, 473-481.
- Rangel, R. M., Amaral Júnior, A. T. D., Gonçalves, L. S. A., Freitas Júnior, S. D. P., & Candido, L. S. (2011). Análise biométrica de ganhos por seleção em população de milho pipoca de quinto ciclo de seleção recorrente. *Revista Ciência Agronômica*, 42, 473-481.
- Rasheed, A., Mahmood, A., Maqbool, R., Albaqami, M., Sher, A., Sattar, A., ... & Wu, Z. (2022). Key insights to develop drought-resilient soybean: A review. *Journal of King Saud University-Science*, 102089.
- Rocha, J. R. D. A. S. D. C., Machado, J. C., & Carneiro, P. C. S. (2018). Multitrait index based on factor analysis and ideotype-design: Proposal and application on elephant grass breeding for bioenergy. *Gcb Bioenergy*, 10(1), 52-60.

- Rocha, J. R. D. A. S. D. C., Nunes, K. V., Carneiro, A. L. N., Marçal, T. D. S., Salvador, F. V., Carneiro, P. C. S., & Carneiro, J. E. S. (2019). Selection of superior inbred progenies toward the common bean ideotype. *Agronomy Journal*, *111*(3), 1181-1189.
- Santchurn, D., Ramdoyal, K., Badaloo, M. G. H., & Labuschagne, M. T. (2014). From sugar industry to cane industry: Evaluation and simultaneous selection of different types of high biomass canes. *Biomass and Bioenergy*, *61*, 82-92.
- Scholander, P. F., Bradstreet, E. D., Hemmingsen, E. A., & Hammel, H. T. (1965). Sap Pressure in Vascular Plants: Negative hydrostatic pressure can be measured in plants. *Science*, *148*(3668), 339-346.
- Seepaul, D. D. R., George, S., Groot, J., & Wright, D. (2019). Drought tolerance classification of common oilseed species using seed germination assay. *Journal of Oilseed Brassica*, *10*(2), 97-105.
- Silva, L. J. D., Medeiros, A. D. D., & Oliveira, A. M. S. (2019). SeedCalc, a new automated R software tool for germination and seedling length data processing. *Journal of Seed Science*, *41*, 250-257.
- Smith, H. F. (1936). A discriminant function for plant selection. *Annals of eugenics*, *7*(3), 240-250.
- Stephens, M. J., Alspach, P. A., Beatson, R. A., Winefield, C., & Buck, E. J. (2012). Genetic parameters and development of a selection index for breeding red raspberries for processing. *Journal of the American Society for Horticultural Science*, *137*(4), 236-242.
- Van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil science society of America journal*, *44*(5), 892-898.
- VILLELA, F. A., DONI FILHO, L., & Sequeira, E. L. (1991). Tabela de Potencial Osmótico em Função da Concentração de Polietileno Glicol 6.000 e da Temperatura.
- Vurukonda, S. S. K. P., Vardharajula, S., Shrivastava, M., & SkZ, A. (2016). Enhancement of drought stress tolerance in crops by plant growth promoting rhizobacteria. *Microbiological research*, *184*, 13-24.
- Wang, X., Li, X., & Dong, S. Screening and Identification of Drought Tolerance of Spring Soybean at Seedling Stage under Climate Change. *Frontiers in Sustainable Food Systems*, 346.
- Woyann, L. G., Meira, D., Matei, G., Zdziarski, A. D., Dallacorte, L. V., Madella, L. A., & Benin, G. (2020). Selection indexes based on linear-bilinear models applied to soybean breeding. *Agronomy Journal*, *112*(1), 175-182.

- Zhang, Y., Liu, Z., Wang, X., Li, Y., Li, Y., Gou, Z., ... & Qiu, L. (2022). Identification of genes for drought resistance and prediction of gene candidates in soybean seedlings based on linkage and association mapping. *The Crop Journal*, *10*(3), 830-839.
- Zuffo, A. M., Steiner, F., Aguilera, J. G., Teodoro, P. E., Teodoro, L. P. R., & Busch, A. (2020). Multi-trait stability index: A tool for simultaneous selection of soya bean genotypes in drought and saline stress. *Journal of Agronomy and Crop Science*, *206*(6), 815-822.

Artigo 2

Espectroscopia NIR para predição do conteúdo de água em folhas de soja e classificação quanto à disponibilidade hídrica

Espectroscopia NIR para predição do conteúdo de água em folhas de soja e classificação quanto à disponibilidade hídrica

Resumo: objetivou-se com este estudo prever o conteúdo de água nas folhas de soja e classificá-las quanto a condição hídrica, por meio de dados NIR e diferentes modelos. Objetivou-se ainda realizar a seleção de variáveis com base na importância das variáveis preditoras determinada pelos modelos Partial Least Squares e Random Forest individualmente e de forma combinada. Para tanto, dois experimentos foram conduzidos, utilizando 100 cultivares comerciais de soja, as quais foram submetidas a duas condições de disponibilidade hídrica (condição controle e condição estresse, sendo este simulado utilizando-se PEG 6000). O estresse por déficit hídrico foi imposto no momento da semeadura e permaneceu por 20 dias. Passados 20 dias, folhas unifolioladas foram coletadas e procedeu-se com a leitura dos espectros em espectrômetro de bancada, e, posteriormente foi realizada a determinação do conteúdo de água, o qual foi realizado pelo método de secagem em estufa. Para realizar a classificação das condições de disponibilidade hídrica e a estimativa do conteúdo de água utilizou-se quatro modelos, sendo eles: Partial Least Squares (PLS), regressão por componentes principais (PCR), Support Vector Machine (SVM) e Random Forest (RF). A fim de reduzir a dimensionalidade dos dados, foi realizada a seleção de variáveis. Para tanto, considerou-se três estratégias: seleção utilizando a importância de variáveis definida com base no modelo PLS, no modelo RF, e em ambos os modelos de forma combinada. Todos os modelos utilizados apresentaram excelente performance para classificar as folhas de soja quanto a disponibilidade hídrica. Todos os modelos apresentaram alta taxa de assertividade tanto na classe da condição controle como na classe da condição estresse. Os modelos PLS e SVM foram os que apresentaram as melhores performances, com acurácia de 99% para ambos os modelos e precisão de 99 e 98%, respectivamente. Para a regressão, os modelos apresentaram bom potencial para realizar a predição do conteúdo de água. As estratégias utilizadas para seleção de variáveis não contribuíram para melhorias significativas no desempenho dos modelos. Dessa forma, pode-se concluir que todos os modelos são eficientes para prever o conteúdo de água em folhas de soja e classificá-las quanto às duas condições de disponibilidade hídrica.

Palavras-chave: *Glycine max* L.; Espectroscopia NIR; déficit hídrico; seleção de variáveis.

1. INTRODUÇÃO

A soja (*Glycine max* L.) apresenta elevada sensibilidade ao deficit hídrico, tanto nos estádios vegetativos, quanto nos estádios reprodutivos. Quando o deficit hídrico acomete a lavoura de soja na fase inicial de desenvolvimento, acarreta sérias consequências para o desenvolvimento da planta, além de contribuir para redução da produção (Du et al., 2020; Wijewardana et al., 2019) e considerável atraso do ciclo da cultura.

A abordagem mais intuitiva para avaliar as relações planta-água é comumente realizada por meio do conteúdo de água da folha (Damm et al., 2018). O conteúdo de água da folha é um indicador importante do status hídrico da planta. Isso significa que essa variável fornece uma medida do deficit hídrico da folha e pode indicar o grau de estresse expresso em condições desfavoráveis como aquelas ocasionadas pela escassez hídrica (Barrs, 1968).

Rallo e Cuevas (2017), relatam que o conteúdo de água tem sido usado como um indicador confiável do bem-estar das plantas e pode ser muito útil para determinar se as plantas estão sofrendo de estresse por deficit hídrico em qualquer um dos estádios fenológicos. Assim sendo, o conteúdo de água foliar pode ser útil para indicar as necessidades hídricas das plantas de soja com o objetivo de reduzir potenciais situações de estresse especialmente nos estádios fenológicos em que a espécie é mais vulnerável a condições extremas.

O deficit hídrico acarreta uma série de modificações a nível, fisiológico, bioquímico, etc., alterando, inclusive, a resposta espectral de plantas (Maimaitiyiming et al., 2017). Folhas de plantas acometidas pelo deficit hídrico, por apresentar menor quantidade de água em seus tecidos, apresentam diferentes comportamentos em determinadas faixas do espectro, apresentando maiores valores de reflectância quando comparadas a folhas de plantas bem irrigadas (Damm et al., 2018). De acordo com Zhang et al. (2019) e Moreira (2012), plantas com maior conteúdo de água em suas folhas refletem a energia de forma mais atenuada. Dessa forma, pode-se associar o comportamento espectral das folhas com variáveis de interesse, como o conteúdo de água foliar.

A espectroscopia de refletância no infravermelho próximo (NIR) é uma ferramenta poderosa para monitorar diversos processos (Zhang et al., 2017). O comportamento espectral de uma amostra constitui uma resposta às suas propriedades, criando-se assim, uma assinatura espectral característica de cada amostra (Pasquini, 2018). Esta abordagem tem sido amplamente aplicada como um método analítico rápido, de fácil uso, já que exige o mínimo ou nenhum preparo de amostra, além de haver a possibilidade de realizar a análise de maneira não destrutiva utilizando-se equipamentos portáteis (Cen et al., 2007). Por exemplo, o conteúdo de água é

tradicionalmente avaliado em laboratório, por meio da secagem das folhas em estufa. Trata-se, portanto, de um método demorado e que demanda muita mão de obra. Neste contexto, o uso da espectroscopia NIR constitui uma ferramenta promissora para determinação do conteúdo de água foliar de forma mais rápida e eficiente.

Na cultura da soja a espectroscopia NIR tem sido especialmente usada para predição do teor de óleo e proteínas (Saito et al., 2021; Leite et al., 2020), monitoramento do teor de ácidos graxos (Zhao et al., 2022), detecção de compostos com isoflavonas e saponinas (Berhow et al., 2020), identificação de sementes quebradas (Wang et al., 2021), classificação de genótipos quanto a disponibilidade hídrica (Crusiol et al., 2021), dentre outras utilidades.

Para o emprego de dados NIR a fim de prever determinada variável de interesse, pode-se empregar diferentes tipos de modelos. Tais modelos utilizam as informações das bandas espectrais mensuradas como variáveis preditoras para a estimativa da variável de interesse. Os modelos Partial Least Squares (PLS) e Regressão por Componentes Principais (PCR), por exemplo, são tradicionalmente usados para este fim (Xu et al., 2020; Borges et al., 2015). Estes possuem alto potencial para lidar com dados de alta dimensionalidade, como ocorre tipicamente na espectroscopia NIR (Pandiselvam et al., 2022; Melado et al., 2021; Li et al., 2020).

Outros modelos como Random Forest (RF), Support Vector Machine (SVM), Redes Neurais Artificiais (RNAs), também podem ser empregados (Wang et al., 2021; Li et al., 2020; Zareef et al., 2020; Santana et al., 2019). De modo geral, algoritmos como estes apresentam uma maior capacidade de capturar padrões complexos e não lineares entre os dados de entrada e saída. Assim, estes modelos têm propiciado melhores performances do que modelos mais tradicionais como PLS e PCR para realizar diferentes tarefas.

Um dos principais desafios ao se trabalhar com espectroscopia NIR, diz respeito a alta dimensionalidade do conjunto de dados. Embora alguns modelos consigam lidar bem com este problema, uma estratégia que vem sendo muito utilizada é a seleção de variáveis. Espera-se com essa abordagem, selecionar comprimentos de onda quimicamente importantes para auxiliar no aprendizado do modelo (Ríos-Reina et al., 2018). Tal estratégia visa melhorar a eficiência computacional, a confiabilidade da predição e a interpretabilidade do modelo (Sem, 2021). Numerosas técnicas de seleção de variáveis podem ser encontradas na literatura tanto para problemas de regressão quanto para a classificação (Yun et al., 2019; Tang et al., 2014; Mehmood et al., 2012; Xiaobo et al., 2010). A predição realizada com as variáveis indicadas pelos modelos PLS e RF já é bem consolidada. Entretanto, a combinação das variáveis selecionadas por cada um deles ainda não foi testada.

Dado o exposto e considerando-se a relevância da cultura da soja a nível mundial, é cada vez mais oportuno que se busque estratégias para solucionar problemas referentes ao déficit hídrico. Dessa forma, objetivou-se prever o conteúdo de água nas folhas de soja e classificá-las quanto a condição hídrica (controle e estresse) por meio de dados NIR e diferentes modelos. Objetivou-se ainda realizar a seleção de variáveis com base na importância das variáveis preditoras determinada pelos modelos PLS e RF individualmente e de forma combinada.

2.MATERIAL E MÉTODOS

Condução dos experimentos e delineamento experimental

Dois experimentos foram conduzidos em casa de vegetação na Universidade Federal de Viçosa (UFV). Em cada experimento foram utilizadas 100 cultivares comerciais de soja, pertencentes ao Banco Ativo de Germoplasma da UFV. Os estudos foram conduzidos sob delineamento de blocos ao acaso, em esquema fatorial com 2 fatores e três repetições. Os fatores foram definidos por 100 cultivares e duas condições de disponibilidade hídrica (controle e estresse). As sementes foram tratadas com Cropstar (tiodicarbe e imidacloprido) + Maxim quatro (azoxistrobina, tiabendazol, fludioxonil e Metalaxil-M) e semeadas em bandejas contendo 15 kg de areia, sendo submetidas, no momento da semeadura a duas condições de disponibilidade hídrica. As condições de disponibilidade hídrica foram: condição controle, na qual a areia foi umedecida em água até que atingisse a capacidade de campo; condição estresse, no potencial osmótico de -0.2 Mpa. Para a condição estresse, uma solução de polietilenoglicol (PEG 6000) foi preparada de acordo com o respectivo potencial osmótico, conforme recomendado por Vilella et al. (1991), sendo esta solução adicionada à areia no momento do plantio.

Após o plantio, as bandejas foram devidamente pesadas, de forma que para cada bandeja foi anotado o seu peso inicial no primeiro dia. A irrigação foi realizada diariamente às 10 e às 16 horas, de acordo com a necessidade, pela diferença de peso das bandejas. Dessa forma, ressalta-se que o estresse por déficit hídrico foi imposto no momento da semeadura e permaneceu por 20 dias após o plantio. O potencial hídrico na folha foi averiguado aos 15 dias após o plantio por meio da bomba de pressão (scholander) (Scholander et al., 1965).

Conteúdo de água e espectroscopia NIR

Passados 20 dias após o plantio, retirou-se duas folhas unifolioladas de cada planta, as quais foram embaladas em papel alumínio, acondicionadas em caixa de isopor e encaminhadas ao laboratório para proceder com as leituras dos respectivos espectros foliares. Somando-se os dois experimentos realizados, foram coletados espectros de 1200 folhas de soja, sendo 600 submetidos à condição controle e 600 submetidos à condição estresse. A aquisição de espectros NIR foi realizada utilizando-se um espectrômetro de bancada (Antaris II FT-NIR Analyzer (Thermo Scientific, EUA)), equipado com um módulo de esfera integrada sendo usado para digitalizar as amostras, as quais foram medidas na faixa espectral de 1000 a 2500 nm. A configuração do equipamento utilizado consistiu em uma média de 32 varreduras por espectro com resolução de 80 cm^{-1} .

Os espectros de refletância foram transformados matematicamente em absorbância por meio do $\log(1/R)$, onde R é a refletância. A temperatura ambiente foi mantida em 25°C , e a umidade permaneceu constante durante todo o processo de aquisição espectral. Posteriormente, avaliou-se o conteúdo de água na folha, o qual foi realizado pelo método tradicional. Essa variável foi mensurada subtraindo-se o peso da massa seca do peso da massa fresca, sendo posteriormente dividido pelo peso da massa fresca.

Modelagem para as tarefas de classificação e regressão

De posse dos dados NIR e conteúdo de água foliar, foram utilizados quatro modelos de aprendizado de máquina para a predição do conteúdo de água (regressão) e para classificar as folhas como pertencentes à condição controle ou estresse (classificação). Para a tarefa de regressão utilizou-se os modelos Partial Least Squares (PLS), regressão por componentes principais (PCR), Support Vector Machine (SVM) e Random Forest (RF). Para a tarefa de classificação, foram empregados os mesmos modelos, com uma variação no caso do PCR, sendo empregado um modelo de classificação baseado em componentes principais (PCC). No PCC, um modelo de regressão logística foi utilizado como substituto para o modelo de regressão linear, permitindo a realização de classificação. A implementação dos modelos foi feita utilizando a linguagem Python, com auxílio da biblioteca Scikit-learn.

Previamente ao treinamento dos modelos, o conjunto de dados foi dividido em subconjuntos de treinamento (70% dos dados disponíveis) e teste (30%). Os dados de treinamento foram usados para o ajuste dos modelos e escolha dos melhores valores de hiperparâmetros e estratégias de pré-processamento dos dados NIR. Para tanto, foi empregada

a validação cruzada do tipo k-fold ($k=5$) com os dados do conjunto de treinamento. Os valores de hiperparâmetros e estratégias de pré-processamento de dados NIR que resultaram nos menores valores de erro na validação cruzada foram escolhidos.

Para a otimização de hiperparâmetros dos modelos de aprendizado de máquina foi empregado o método grid-search. Para o PLS e PCR/PCC foram ajustados o número de variáveis latentes/componentes principais utilizados, sendo avaliado o uso de 4 a 40 variáveis latentes, com intervalo de 2. Para o modelo SVM, ajustou-se o valor do hiperparâmetro de regularização C, buscando entre 50 a 500, com intervalo de 50. Para o modelo RF, ajustou-se a profundidade máxima das árvores de decisão (5, 7, 9 e 11) e o número variáveis consideradas para a divisão dos nós em cada árvore de decisão (100% e 33% das variáveis).

Para o uso de dados NIR, a utilização de técnicas de pré-processamento é de grande importância (Pasquini, 2018). Portanto, foram avaliados o uso dos dados brutos e pré-processados utilizando os métodos SNV (Standard Normal Variate) e Savitzky-Golay (SG). Foram consideradas diferentes configurações do SG, variando-se a largura da janela (w) e ordem da derivada (d): SG1 ($w=7$, $d=1$), SG2 ($w=11$, $d=1$), SG3 ($w=7$, $d=2$) e SG4 ($w=11$, $d=2$). O grau do polinômio utilizado para a suavização foi igual a 2. Por fim, foram testadas 10 estratégias de pré-processamento: dados brutos, SNV, SG1, SG2, SG3, SG4, SNV+SG1, SNV+SG2, SNV+SG3 e SNV+SG4.

Seleção de variáveis

A fim de reduzir a alta dimensionalidade dos dados (1557 variáveis), foi também avaliado o uso de seleção de variáveis. Esta técnica consiste na seleção das variáveis preditoras de maior importância para o problema de interesse, podendo levar a obtenção de ganhos de performance. No presente estudo, a seleção de variáveis foi realizada com base na importância das variáveis preditoras obtida com base nos modelos PLS e RF gerados com todo o conjunto de variáveis. Após ordenar as variáveis de acordo com sua importância, adotou-se como critério de seleção a escolha das 300 variáveis de maior importância para o modelo considerado. Para a seleção de variáveis, considerou-se três estratégias: seleção utilizando a importância de variáveis definida com base no modelo PLS, com base no modelo RF, e com base em ambos os modelos de forma combinada. Após a seleção de variáveis, todos os modelos empregados neste estudo (PLS, PCR/PCC, SVM e RF) foram ajustados novamente.

Critérios para avaliação da performance dos modelos

Os modelos foram avaliados utilizando os dados do conjunto de teste. No caso dos modelos de regressão, foram utilizados os indicadores estatísticos raiz do erro quadrático médio (RMSE), coeficiente de determinação (R^2), viés médio (MBE) e erro absoluto médio (MAE), conforme equações abaixo.

$$RMSE = \sqrt{\frac{1}{n} \sum (P_i - O_i)^2}$$

$$MAE = \frac{1}{n} \sum |P_i - O_i|$$

$$MBE = \frac{1}{n} \sum (P_i - O_i)$$

$$R^2 = \left[\frac{\sum (P_i - \bar{P})(O_i - \bar{O})}{\sqrt{(\sum (P_i - \bar{P})^2)(\sum (O_i - \bar{O})^2)}} \right]^2$$

Onde, P_i é o valor predito, O_i é o valor observado, \bar{P} é a média dos valores preditos, \bar{O} é a média dos valores observados e n é o número de pares de dados.

Para os modelos utilizados para classificação, foram utilizados os indicadores acurácia, precisão, recall e F1-score, conforme as equações apresentadas a seguir.

$$\text{Acurácia} = \frac{VP+VN}{VP+VN+FP+FN}$$

$$\text{Precisão} = \frac{VP}{VP+FP}$$

$$\text{Recall} = \frac{VP}{VP+FN}$$

$$\text{F1 score} = 2 * \frac{\text{Precisão} * \text{Recall}}{\text{Precisão} + \text{Recall}}$$

Onde, VP é verdadeiro positivo, VN é verdadeiro negativo, FP e falso positivo e FN é falso negativo.

Visão geral da estrutura metodológica empregada

Todas as etapas envolvidas na coleta de dados, bem como nas modelagens com base nos dados NIR obtidos, são sumarizadas na Figura 1.

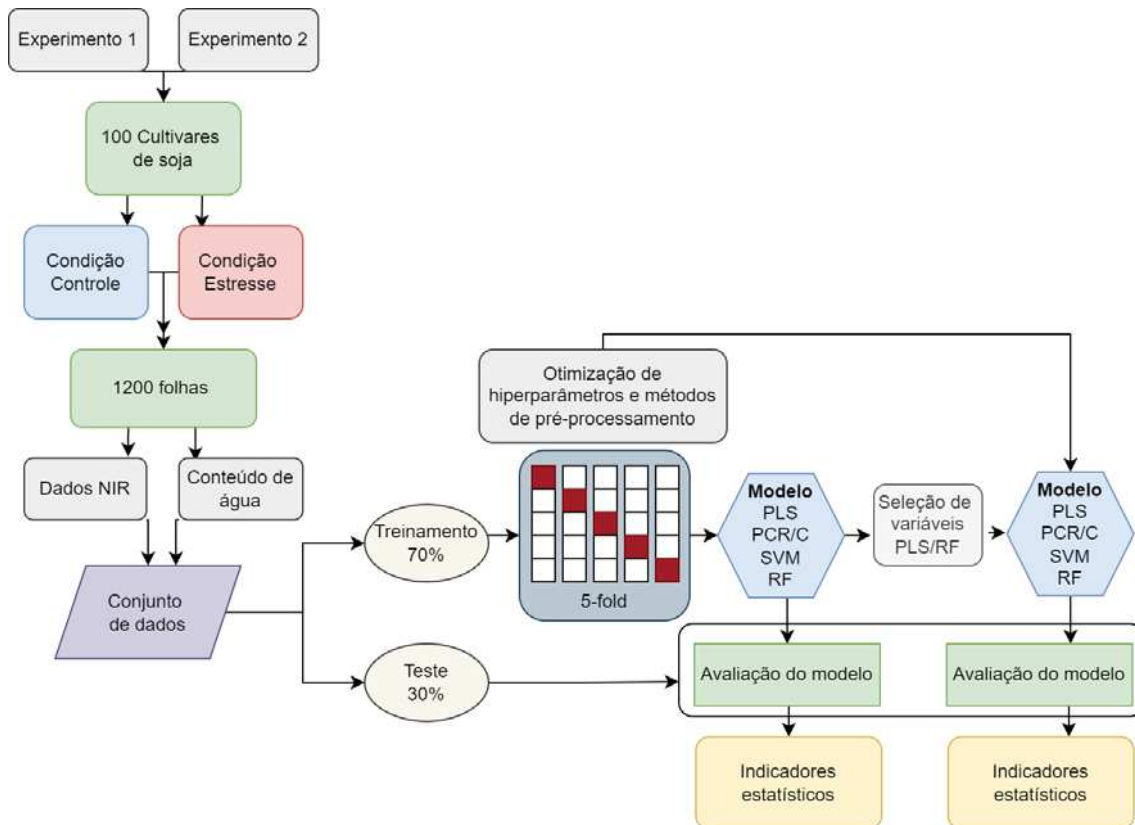


Figura 1. Esquema geral dos principais procedimentos realizados no presente estudo.

3. RESULTADOS E DISCUSSÃO

Análise exploratória dos dados

Nos dois experimentos realizados, o conteúdo de água na folha foi superior na condição controle, quando comparada à condição estresse, sendo tal diferença ligeiramente mais acentuada no segundo experimento (Figura 2). Para a condição controle, a maior parte dos valores de conteúdo de água encontra-se variando de cerca de 70 a 82%. Já para a condição estresse, os valores encontram-se distribuídos entre cerca de 73 a 87%.

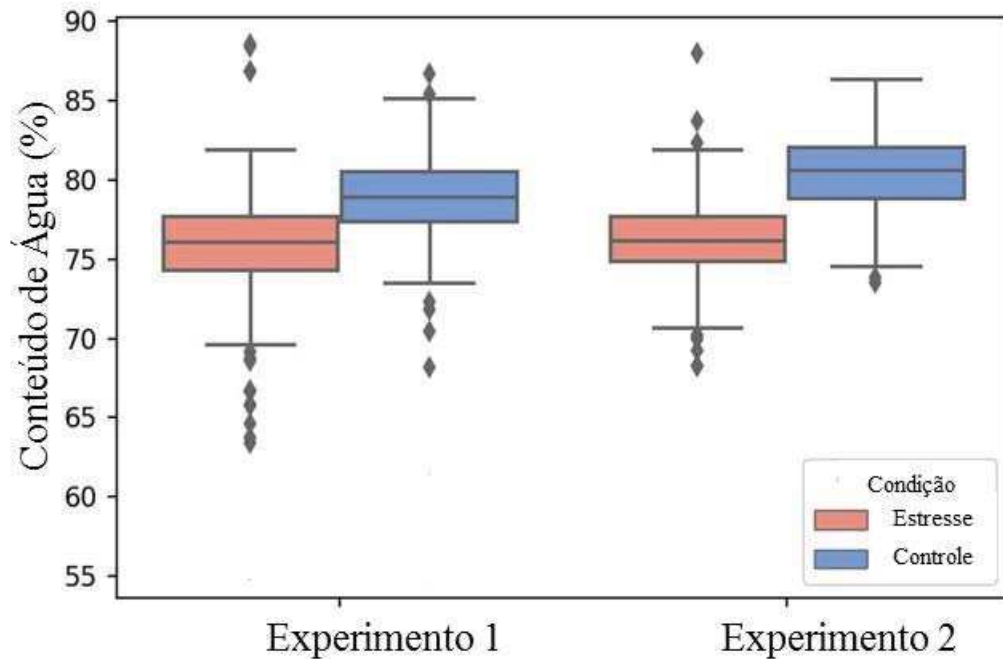


Figura 2. Boxplot com os valores médios observados para a variável conteúdo de água em folhas de soja submetidas a duas condições de disponibilidade hídrica.

A reflectância média nos comprimentos de onda considerados para a condição controle e a condição estresse, encontra-se exposta na Figura 3. Observou-se diferenças no comportamento espectral das folhas de soja submetidas à condição estresse em relação a condição controle. Braga et al. (2021), caracterizando o conteúdo de água em folhas de soja submetidas às condições de irrigação e deficit hídrico, observaram maiores valores de refletância para plantas com deficit hídrico. No presente estudo, as diferenças nas reflectâncias encontram-se mais evidentes em comprimentos de ondas específicos. Na região que compreende os comprimentos de onda de 1400 a pouco mais de 1800nm, as folhas de soja apresentaram maiores diferenças na reflectância ao se comparar a condição controle e a condição estresse.

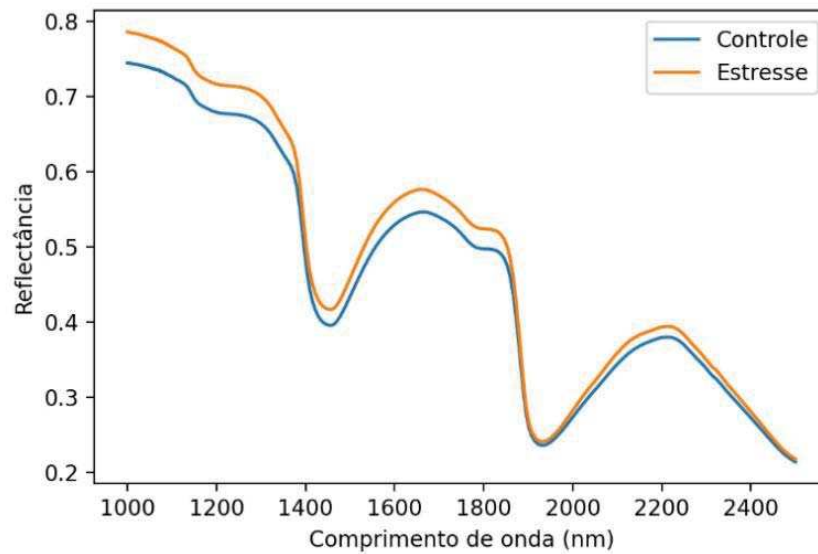


Figura 3. Valores médios de reflectância de folhas de soja submetidas a duas condições de disponibilidade hídrica.

Ressalta-se que as leituras espectrais obtidas neste estudo, compreendem os comprimentos de onda referentes a parte da região do infravermelho próximo, a qual encontra-se entre 700 a 1300nm, e parte do infravermelho médio, que está entre 1300 a 3000nm. Os picos observados neste estudo nos valores de reflectância encontram-se justamente na faixa do infravermelho médio, a qual apresenta relação com o conteúdo de água foliar. Nesta região, a água apresenta alta capacidade de absorção da radiação eletromagnética, contribuindo para redução da reflectância das folhas com maiores conteúdos de água (Moreira, 2012). Na região espectral do infravermelho médio, as bandas de maior absorção de água ocorrem em 1400, 1900 e 2500nm, como pode ser verificado no presente trabalho. Braga et al. (2021) também observaram este mesmo comportamento espectral ao analisar o conteúdo de água em 10 cultivares de soja submetidas a condição controle e a condição de deficit hídrico.

De modo geral, quando a planta é acometida pelo deficit hídrico, ocorre em seu interior uma série de reações bioquímicas na tentativa de contornar essa situação. Uma das primeiras consequências do deficit hídrico é o fechamento dos estômatos. Em decorrência disso, a planta perde menos água, e consequentemente, reduz-se a taxa fotossintética, levando ao aumento da concentração interna de CO_2 , já que a planta prioriza o processo respiratório (Marinho et al., 2016). Associado a isto, do ponto de vista do comportamento espectral, plantas com menor conteúdo de água nas folhas refletem energia de forma mais acentuada (Moreira, 2012).

Ressalta-se, portanto, que as diferenças na reflectância ao longo do espectro desempenham um papel importante no monitoramento e na diferenciação do status hídrico da

cultura da soja e podem contribuir para o gerenciamento da irrigação e tomadas de decisões ao longo do ciclo da cultura. Ademais, o dinamismo da refletância espectral sob diferentes condições de disponibilidade hídrica, possibilita o desenvolvimento de modelos baseados em espectroscopia para estimativa de conteúdo de água em folhas de soja (Das et al., 2021).

Modelagem para as tarefas de classificação e regressão

Nas Tabelas 1 e 2 encontram-se expostas as melhores estratégias de pré-processamento, bem como os melhores valores dos hiperparâmetros testados, para todos os modelos utilizados.

Tabela 1. Melhores estratégias de pré-processamento obtidas para cada modelo/abordem utilizadas.

Modelo	Pré-processamentos			
	Regressão		Classificação	
	Antes SV	Após SV	Antes SV	Após SV
PLS	Brutos	Brutos	Brutos	Brutos
PCR	Brutos	Brutos	SNV	SNV
SVM	SG1	SG1	SNV+SG2	SNV+SG2
RF	SNV+SG2	SNV+SG2	SNV+SG1	SNV+SG1

*SV = Seleção de variáveis; SNV = Standard Normal Variate; SG = Savitzky-Golay.

SG1 (w=7, d=1), SG2 (w=11, d=1), SG3 (w=7, d=2) e SG4 (w=11, d=2); w = largura da janela e d = ordem da derivada.

Tabela 2. Otimização dos hiperparâmetros utilizados para todos os modelos utilizados para realizar as tarefas de classificação e regressão, antes e após a seleção de variáveis.

Modelo	Hiperparâmetro	Otimização de hiperparâmetros							
		Regressão				Classificação			
		Antes SV	SV-PLS	SV-RF	SV-PLS/RF	Antes SV	SV-PLS	SV-RF	SV-PLS/RF
PLS	n°var. latentes	14	16	16	16	14	12	14	16
PCR	n°componentes	36	40	34	38	16	14	16	12
SVM	C	100	450	500	500	400	500	500	500
RF	Prof. árvore	11	11	11	9	7	9	9	7
	Máx. var.	0.33	0.33	1	0.33	0.33	0.33	0.33	0.33

SV= seleção de variáveis; Prof. Árvore= profundidade da árvore; Máx. var.= número máximo de variáveis para cada árvore.

Classificação quanto à condição controle e condição estresse

De modo geral, todos os modelos empregados foram eficientes na classificação quanto às duas condições de disponibilidade hídrica (Tabela 3). Os modelos PLS e SVM foram os que apresentaram as melhores performances, com acurácia de 99% para ambos os modelos e precisão de 99 e 98%, respectivamente. Isso significa que os números de falsos positivos foram menores nesses modelos. Estes dois modelos também apresentaram maiores valores para os indicadores Recall e F1 score.

Tabela 3. Indicadores estatísticos para os quatro modelos utilizados para classificação de folhas de soja quanto à condição de disponibilidade hídrica.

Modelo	Acurácia	Precisão	recall	f1 score
PLS	0,99	0,99	0,99	0,99
PCC	0,92	0,93	0,92	0,92
SVM	0,99	0,98	0,99	0,99
RF	0,96	0,96	0,96	0,96

Por se tratar de um dos métodos supervisionados mais amplamente usados para classificação (Yi et al., 2016) em estudos de espectroscopia NIR, já era de se esperar que o modelo PLS apresentasse um bom desempenho. Outros estudos demonstraram a eficácia deste modelo para fins de classificação usando dados de espectroscopia NIR (Mansouri et al., 2021; Monago-Maraña et al., 2021; Borraz-Martínez, et al., 2019; Górski et al., 2016).

Com relação a eficiência dos modelos SVM e RF, um estudo recente avaliou a performance desses modelos ao classificar sementes de soja quebradas usando espectroscopia NIR (Wang et al., 2021). Os autores observaram bons resultados, com precisão de 72 e 80% para os modelos SVM e RF, respectivamente. De acordo com Zareef et al. (2020), o modelo SVM, por apresentar boa capacidade para lidar com dados de alta dimensionalidade, é uma ferramenta eficaz para modelagens envolvendo dados NIR, os quais tipicamente apresentam alta dimensionalidade.

Diferentemente do observado nos demais modelos testados, o modelo PCC apresentou a pior performance ao classificar a condição controle e a condição estresse. Isso provavelmente ocorre porque ao gerar as variáveis latentes (componentes principais) utilizadas pelo modelo de predição, não são utilizados dados da variável resposta (Neto et al., 2021). De maneira contrária, o PLS cria variáveis latentes de forma a priorizar a capacidade do modelo em prever a variável resposta, conduzindo a obtenção de melhores resultados.

Com base na matriz de confusão apresentada na Figura 4, verifica-se mais uma vez o alto desempenho dos modelos utilizados no presente estudo. Todos os modelos apresentaram alta taxa de assertividade tanto na classe da condição controle como na classe da condição estresse. Ressalta-se ainda a elevada performance do modelo PLS, o qual apresentou apenas um erro na condição estresse e dois erros na condição controle.

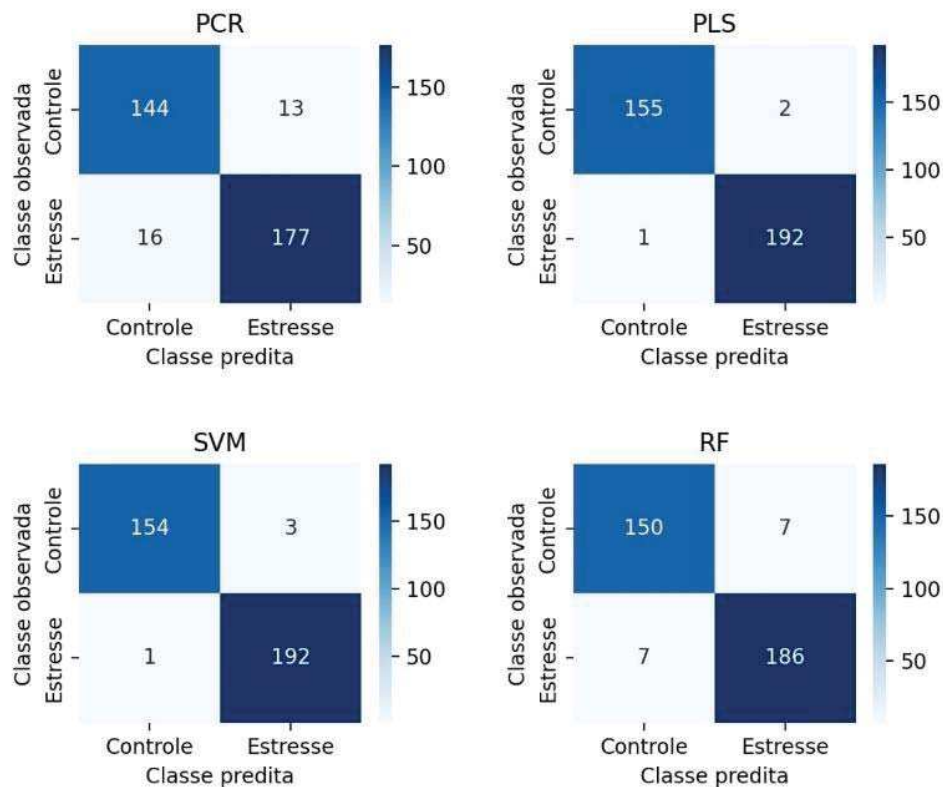


Figura 4. Matriz de confusão com as classes observadas e previstas de folhas de soja quanto às duas condições de disponibilidade hídrica para os modelos utilizados.

Predição do conteúdo de água em folhas de soja

Embora tenham apresentado valores de R^2 relativamente baixos, todos os modelos empregados neste estudo apresentaram bom potencial para prever o conteúdo de água em folhas de soja (Tabela 4).

Ao analisar os indicadores estatísticos, verifica-se que os erros de predição foram baixos, o que indica que os modelos são capazes de realizar as estimativas com êxito. Os modelos mais tradicionais (PLS e PCR) apresentaram resultados ligeiramente melhores, já que os valores de MAE obtidos para tais modelos (1,46 e 1,45, respectivamente) foram inferiores aos obtidos para os modelos SVM e RF (1,51 e 1,55, respectivamente). Quanto aos valores de

RMSE e MBE, de modo geral, valores similares foram observados, com exceção para o modelo SVM, o qual obteve valor de MBE ligeiramente superior aos demais.

Tabela 4. Indicadores estatísticos para os modelos utilizados para prever o conteúdo de água de folhas de soja submetidas à duas condições de disponibilidade hídrica.

Modelo	RMSE	MBE	MAE	R²
PLS	2,53	-0,02	1,46	0,51
PCR	2,54	-0,01	1,45	0,50
SVM	2,51	0,14	1,51	0,51
RF	2,52	-0,04	1,55	0,51

Os resultados apresentados no presente estudo, para a estimativa do conteúdo de água diferem dos resultados encontrados em um estudo recente realizado por Das et al. (2021). Os autores utilizaram diferentes modelos, dentre eles o PLS e o RF para realizar a predição do conteúdo de água e observaram R² de 0,94 e 0,88 para ambos os modelos, respectivamente, na etapa de validação. Entretanto, os autores observaram maiores valores para RMSE e MAE (5,49 e 1,06; 7,5 e 0,18), respectivamente para os modelos PLS e RF. Outros autores também obtiveram bons resultados ao utilizar o modelo PLS para prever o conteúdo de água (Das et al., 2017; Mirzaie et al., 2014; Ullah et al., 2014), utilizando abordagens diferentes das empregadas neste estudo.

Na Figura 5, encontram-se expostos os valores observados e estimados de conteúdo de água em folhas de soja. De maneira geral, verifica-se que os modelos apresentaram boa capacidade de predição, uma vez que a maior parte dos valores preditos encontra-se próxima a linha 1:1. Apesar dos modelos terem apresentado comportamento similar, observa-se que os modelos RF e SVM aparentam ligeira tendência em superestimar valores mais baixos de conteúdo de água e subestimar valores mais elevados de conteúdo de água.

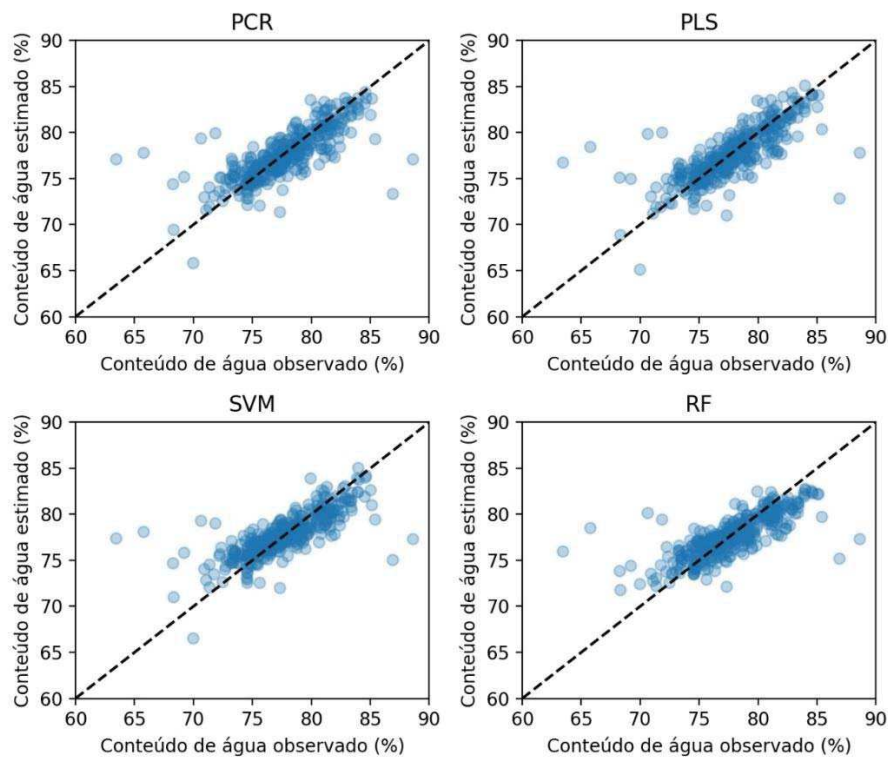


Figura 5. Valores observados e estimados para a variável conteúdo de água de folhas de soja submetidas a duas condições de disponibilidade hídrica para cada um dos modelos utilizados.

Seleção de variáveis

A análise da importância das variáveis (comprimentos de onda) obtida pelos modelos PLS e RF indicou diferenças nos comportamentos dos modelos (Figura 6). Para o modelo PLS, observou-se maior concentração das bandas mais importantes em regiões mais específicas, tanto para classificação quanto para regressão. De forma contrária, o modelo RF indicou, como mais importantes, bandas espalhadas de forma mais distribuída ao longo dos comprimentos de onda considerados. Ressalta-se que o RF considerou algumas bandas da região situada a partir de 1700 nm como sendo destacadamente superiores às demais bandas selecionadas.

Ao relacionar a seleção de variáveis com a faixa espectral do infravermelho médio (região acima de 1300 nm) a qual apresenta elevada relação com a absorção de água, verifica-se que os modelos não indicaram apenas estas bandas como as mais importantes. O modelo PLS indicou um número considerável de bandas nas faixas espectrais próximo a 1400, 1900 e 2500 nm tanto para a tarefa de regressão e quanto para classificação. Já o modelo RF indicou bandas parecidas para realizar a tarefa de classificação.

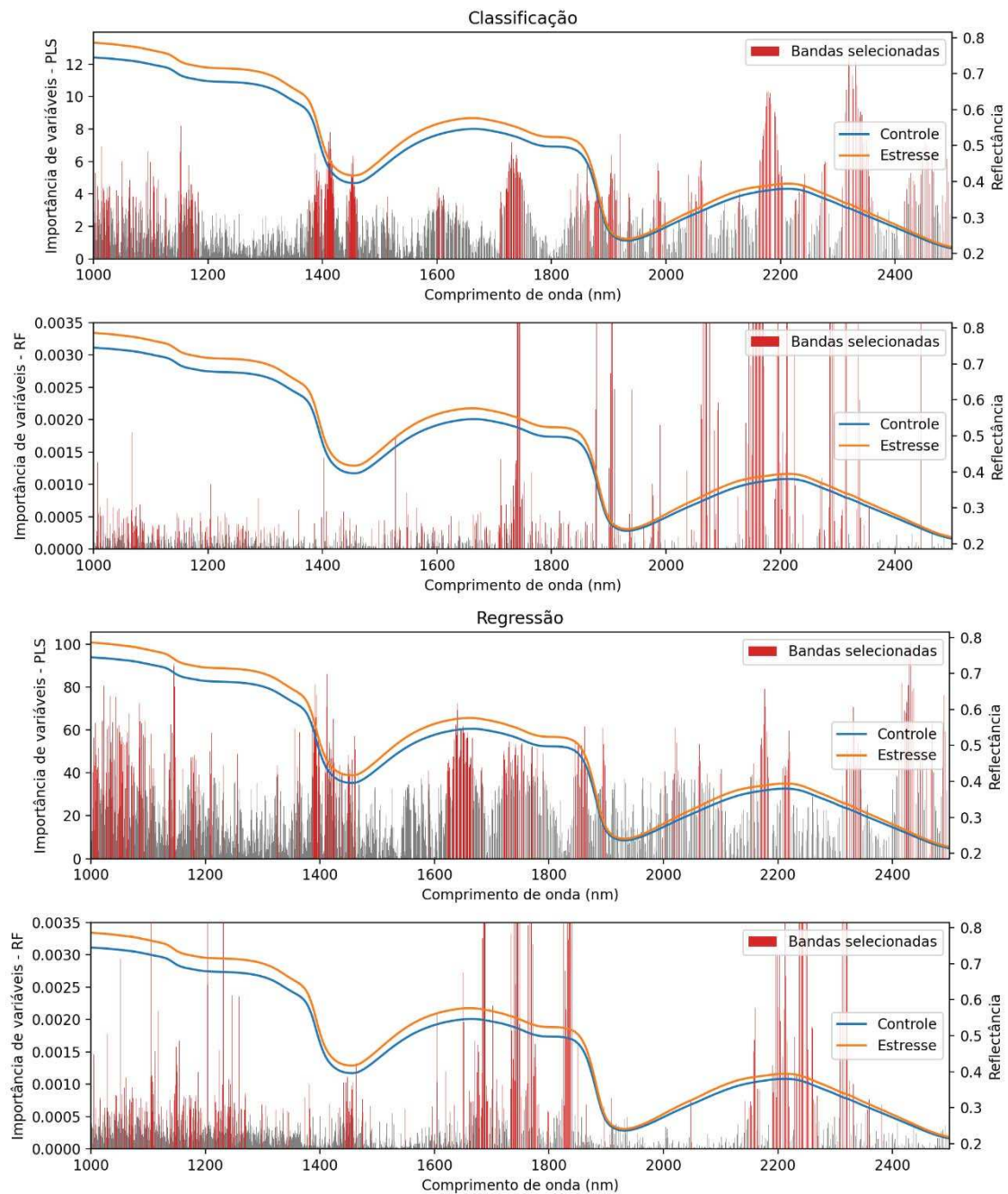


Figura 6. Importância de variáveis indicadas pelos modelos PLS e RF nas tarefas de classificação e regressão. Bandas destacadas em vermelho foram selecionadas. As linhas azuis e laranja indicam os valores médios de reflectância de folhas de soja submetidas a duas condições de disponibilidade hídrica.

A seleção de variáveis realizada com as diferentes estratégias avaliadas não resultou em ganho de desempenho para os modelos testados, tanto para classificação quanto para regressão (Tabela 5). Além disso, com exceção do modelo PLS, utilizado para classificação, todos os demais modelos apresentaram performances inferiores (Tabela 3 e 4).

Tabela 5. Indicadores estatísticos para as estratégias de seleção de variáveis para classificação de folhas de soja quanto a disponibilidade hídrica e estimativa do conteúdo de água.

Classificação					Regressão				
PLS									
Modelo	Acurácia	Precisão	Recall	F1*	Modelo	RMSE	MBE	MAE	R ²
PLS	0,99	0,99	0,99	0,99	PLS	2,68	-0,01	1,57	0,46
PCC	0,87	0,92	0,84	0,88	PCR	2,63	-0,06	1,57	0,47
SVM	0,97	0,97	0,97	0,97	SVM	2,61	0,18	1,68	0,50
RF	0,94	0,95	0,94	0,94	RF	2,67	0,06	1,71	0,45
RF									
Modelo	Acurácia	Precisão	Recall	F1*	Modelo	RMSE	MBE	MAE	R ²
PLS	0,99	0,99	0,98	0,99	PLS	2,90	-0,08	1,85	0,39
PCC	0,88	0,92	0,86	0,89	PCR	2,63	-0,04	1,56	0,46
SVM	0,97	0,97	0,97	0,97	SVM	2,56	0,20	1,58	0,50
RF	0,94	0,95	0,94	0,94	RF	2,57	0,04	1,57	0,49
PLS/RF									
Modelo	Acurácia	Precisão	Recall	F1*	Modelo	RMSE	MBE	MAE	R ²
PLS	0,99	0,99	0,99	0,99	PLS	2,72	0,00	1,65	0,45
PCC	0,88	0,93	0,85	0,89	PCR	2,60	-0,02	1,54	0,48
SVM	0,97	0,97	0,98	0,97	SVM	2,59	0,19	1,64	0,50
RF	0,95	0,97	0,93	0,95	RF	2,65	0,04	1,68	0,45

*F1= F1 score

Para classificação, todas as estratégias de seleção de variáveis avaliadas apresentaram performances similares. Entretanto, verificou-se que a seleção de variáveis com base nos modelos PLS e RF de forma conjunta apresentou resultado ligeiramente melhor do que o uso individual dos modelos. Já para a regressão, a seleção de variáveis baseada no modelo RF e RF combinado com PLS resultaram em desempenho pouco melhor do que o uso da seleção de variáveis baseada apenas no uso do PLS.

A seleção de variáveis comumente contribui com melhores ganhos em estudos de modelagem envolvendo a espectroscopia NIR (Yun et al., 2019; Ríos-Reina et al., 2018; Tang et al., 2014). Entretanto, nem sempre a aplicação dessas técnicas realmente resulta em ganhos de desempenho, conforme observado no presente estudo. Ferreira e Peternelli (2021) avaliou diferentes métodos de seleção de variáveis em dois conjuntos de dados NIR e também verificou reduções de performance em relação aos modelos com dados completos. De forma contrária, Sem (2021), utilizou diferentes métodos de seleção de variáveis e demonstrou que a implementação destas abordagens no processo de construção do modelo pode melhorar consideravelmente a tarefa de classificação. Trabalhos futuros podem ser realizados a fim de avaliar outras estratégias para seleção de variáveis.

4. CONCLUSÕES

Todos os modelos empregados são eficientes para realizar a classificação de folhas de soja quanto às duas condições de disponibilidade hídrica avaliadas neste estudo. De forma semelhante, todos os modelos empregados para prever o conteúdo de água em folhas de soja também apresentaram bom desempenho.

O modelo PLS e SVM apresentaram os melhores resultados para a tarefa de classificação. Para a regressão, os modelos PLS e PCR apresentaram as melhores performances.

A seleção de variáveis, com todas as estratégias avaliadas, não contribuiu para a melhoria da performance de nenhum dos modelos utilizados.

5. REFERÊNCIAS

- Barrs, H. D. (1968). Determination of water deficits in plant tissues. *Water deficits and plant growth*, Vol. 1, 235-368.
- Berhow, M. A., Singh, M., Bowman, M. J., Price, N. P., Vaughn, S. F., & Liu, S. X. (2020). Quantitative NIR determination of isoflavone and saponin content of ground soybeans. *Food chemistry*, 317, 126373.
- Borges, G. R., Farias, G. B., Braz, T. M., Santos, L. M., Amaral, M. J., Fortuny, M., ... & Santos, A. F. (2015). Use of near infrared for evaluation of droplet size distribution and water content in water-in-crude oil emulsions in pressurized pipeline. *Fuel*, 147, 43-52.
- Borraz-Martínez, S., Simó, J., Gras, A., Mestre, M., & Boqué, R. (2019). Multivariate Classification of *Prunus dulcis* Varieties using Leaves of Nursery Plants and Near-Infrared Spectroscopy. *Scientific reports*, 9(1), 1-9.
- Braga, P., Crusiol, L. G. T., Nanni, M. R., Caranhato, A. L. H., Fuhrmann, M. B., Nepomuceno, A. L., ... & Mertz-Henning, L. M. (2021). Vegetation indices and NIR-SWIR spectral bands as a phenotyping tool for water status determination in soybean. *Precision Agriculture*, 22(1), 249-266.
- Cen, H., & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology*, 18(2), 72-83.
- Crusiol, L. G. T.; Braga, P.; Nanni, R. M.; Furlanetto, R. H., Sibaldelli, R. N. R., Cezar, E., ... & Bouças Farias, J. R. (2021). Using leaf-based hyperspectral reflectance for genotype classification within a soybean germplasm collection assessed under different levels of water availability. *International Journal of Remote Sensing*, 42(21), 8196-8215.
- Das, B., Sahoo, R. N., Pargal, S., Krishna, G., Verma, R., Viswanathan, C., ... & Gupta, V. K. (2021). Evaluation of different water absorption bands, indices and multivariate models for water-deficit stress monitoring in rice using visible-near infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 247, 119104.
- Das, B., Sahoo, R. N., Pargal, S., Krishna, G., Verma, R., Chinnusamy, V., ... & Gupta, V. K. (2017). Comparison of different uni-and multi-variate techniques for monitoring leaf water status as an indicator of water-deficit stress in wheat through spectroscopy. *Biosystems Engineering*, 160, 69-83.

- Ferreira, R. A., & Peternelli, L. A. Comparison of variable selection methods in predictive models applied to near-infrared and genomic data. *Genetics and Molecular Research* 20 (2), 18909.
- Górski, Ł., Sordoń, W., Ciepela, F., Kubiak, W. W., & Jakubowska, M. (2016). Voltammetric classification of ciders with PLS-DA. *Talanta*, 146, 231-236.
- Leite, D. C., Corrêa, A. A. P., Júnior, L. C. C., de Lima, K. M. G., de Moraes, C. D. L. M., Vianna, V. F., ... & Unêda-Trevisoli, S. H. (2020). Non-destructive genotypes classification and oil content prediction using near-infrared spectroscopy and chemometric tools in soybean breeding program. *Journal of Food Composition and Analysis*, 91, 103536.
- Li, X., Zhang, L., Zhang, Y., Wang, D., Wang, X., Yu, L., ... & Li, P. (2020). Review of NIR spectroscopy methods for nondestructive quality analysis of oilseeds and edible oils. *Trends in Food Science & Technology*, 101, 172-181.
- Maimaitiyiming, M., Ghulam, A., Bozzolo, A., Wilkins, J. L., & Kwasniewski, M. T. (2017). Early detection of plant physiological responses to different levels of water stress using reflectance spectroscopy. *Remote Sensing*, 9(7), 745.
- Mansouri, M. A., Ziemons, E., Sacré, P. Y., Kharbach, M., Barra, I., Cherrah, Y., ... & Bouklouze, A. (2021). Classification of polymorphic forms of fluconazole in pharmaceuticals by FT-IR and FT-NIR spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis*, 196, 113922.
- Marinho, J. P., Kanamori, N., Ferreira, L. C., Fuganti-Pagliarini, R., Carvalho, J. D. F. C., Freitas, R. A., ... & Nepomuceno, A. L. (2016). Characterization of molecular and physiological responses under water deficit of genetically modified soybean plants overexpressing the AtAREB1 transcription factor. *Plant molecular biology reporter*, 34(2), 410-426.
- Mehmood, T., Liland, K. H., Snipen, L., & Sæbø, S. (2012). A review of variable selection methods in partial least squares regression. *Chemometrics and intelligent laboratory systems*, 118, 62-69.
- Melado-Herreros, A., Nieto-Ortega, S., Olabarrieta, I., Gutiérrez, M., Villar, A., Zufía, J., ... & Roger, J. M. (2021). Postharvest ripeness assessment of 'Hass' avocado based on development of a new ripening index and Vis-NIR spectroscopy. *Postharvest Biology and Technology*, 181, 111683.

- Monago-Maraña, O., Eskildsen, C. E., Galeano-Díaz, T., de la Peña, A. M., & Wold, J. P. (2021). Untargeted classification for paprika powder authentication using visible–Near infrared spectroscopy (VIS-NIRS). *Food Control*, *121*, 107564.
- Moreira, M. A. Fundamentos do sensoriamento remoto e metodologias de aplicação. 4 ed. Viçosa – MG, 2011.
- Neto, A. J. S., Toledo, J. V., Zolnier, S., Lopes, D. D. C., Pires, C. V., & da Silva, T. G. F. (2017). Prediction of mineral contents in sugarcane cultivated under saline conditions based on stalk scanning by Vis/NIR spectral reflectance. *biosystems engineering*, *156*, 17-26.
- Pandiselvam, R., Mahanti, N. K., Manikantan, M. R., Kothakota, A., Chakraborty, S. K., Ramesh, S. V., & Beegum, P. S. (2022). Rapid detection of adulteration in desiccated coconut powder: vis-NIR spectroscopy and chemometric approach. *Food Control*, *133*, 108588.
- Pasquini, C. (2018). Near infrared spectroscopy: A mature analytical technique with new perspectives—A review. *Analytica chimica acta*, *1026*, 8-36.
- Ríos-Reina, R., García-González, D. L., Callejón, R. M., & Amigo, J. M. (2018). NIR spectroscopy and chemometrics for the typification of Spanish wine vinegars with a protected designation of origin. *Food Control*, *89*, 108-116.
- Saito, Y., Itakura, K., Kuramoto, M., Kaho, T., Ohtake, N., Hasegawa, H., ... & Kondo, N. (2021). Prediction of protein and oil contents in soybeans using fluorescence excitation emission matrix. *Food Chemistry*, *365*, 130403.
- Santana, F. B., Neto, W. B., & Poppi, R. J. (2019). Random forest as one-class classifier and infrared spectroscopy for food adulteration detection. *Food chemistry*, *293*, 323-332.
- Sem, V. (2021). Interpretability of selected variables and performance comparison of variable selection methods in a polyethylene and polypropylene NIR classification task. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *258*, 119850.
- Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. *Data classification: Algorithms and applications*, 37.
- Ullah, S., Skidmore, A. K., Ramoelo, A., Groen, T. A., Naeem, M., & Ali, A. (2014). Retrieval of leaf water content spanning the visible to thermal infrared spectra. *ISPRS journal of photogrammetry and remote sensing*, *93*, 56-64.

- Wang, L., Huang, Z., & Wang, R. (2021). Discrimination of cracked soybean seeds by near-infrared spectroscopy and random forest variable selection. *Infrared Physics & Technology*, *115*, 103731.
- Wijewardana, C., Reddy, K. R., & Bellaloui, N. (2019). Soybean seed physiology, quality, and chemical composition under soil moisture stress. *Food Chemistry*, *278*, 92-100.
- Xiaobo, Z., Jiewen, Z., Povey, M. J., Holmes, M., & Hanpin, M. (2010). Variables selection methods in near-infrared spectroscopy. *Analytica chimica acta*, *667*(1-2), 14-32.
- Yi, L., Dong, N., Yun, Y., Deng, B., Ren, D., Liu, S., & Liang, Y. (2016). Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. *Analytica chimica acta*, *914*, 17-34.
- Yun, Y. H., Li, H. D., Deng, B. C., & Cao, D. S. (2019). An overview of variable selection methods in multivariate analysis of near-infrared spectra. *TrAC Trends in Analytical Chemistry*, *113*, 102-115.
- Zareef, M., Chen, Q., Hassan, M. M., Arslan, M., Hashim, M. M., Ahmad, W., ... & Agyekum, A. A. (2020). An overview on the applications of typical non-linear algorithms coupled with NIR spectroscopy in food analysis. *Food Engineering Reviews*, 1-18.
- Zhang, Y., Luo, L., Li, J., Li, S., Qu, W., Ma, H., ... & Ye, X. (2017). In-situ and real-time monitoring of enzymatic process of wheat gluten by miniature fiber NIR spectrometer. *Food Research International*, *99*, 147-154.
- Zhang, Z., Tang, B. H., & Li, Z. L. (2019). Retrieval of leaf water content from remotely sensed data using a vegetation index model constructed with shortwave infrared reflectances. *International Journal of Remote Sensing*, *40*(5-6), 2313-2323.
- Zhao, L., Zhang, M., Wang, H., & Mujumdar, A. S. (2022). Monitoring of free fatty acid content in mixed frying oils by means of LF-NMR and NIR combined with BP-ANN. *Food Control*, *133*, 108599.

CONCLUSÕES GERAIS

Os dois índices de seleção FAI-BLUP e MGIDI utilizados neste estudo possibilitaram a seleção de 15 cultivares de soja. Foram selecionadas 12 cultivares comuns aos dois índices.

As cultivares M 9144 RR, BMX TITAN RR foram as que mais se aproximaram do ideótipo. As cultivares de soja aqui selecionadas podem ser utilizadas como fonte de germoplasma em estudos futuros em programas de melhoramento visando a tolerância ao déficit hídrico.

Todos os modelos de machine learning empregados foram eficientes para realizar a classificação das folhas de soja quanto à condição de disponibilidade hídrica.

Todos os modelos empregados para prever o teor de água em folhas de soja, apresentaram bom desempenho. Os modelos PLS e SVM apresentaram os melhores resultados para a tarefa de classificação. Para regressão, os modelos PLS e PCR apresentaram os melhores desempenhos.