

Universidade de São Paulo  
Escola Superior de Agricultura “Luiz de Queiroz”

Utilização de dados hiperespectrais para diferenciação de variedades de cana-  
de-açúcar (*Saccharum officinarum* L.)

**Luiz Augusto Leite Ribeiro**

Dissertação apresentada para obtenção do título de  
Mestre em Ciências. Área de concentração: Engenharia  
de Sistemas Agrícolas

Piracicaba  
2022

Luiz Augusto Leite Ribeiro  
Engenheiro Agrônomo

Utilização de dados hiperespectrais para diferenciação de variedades de cana-de-açúcar  
(*Saccharum officinarum* L.)

versão revisada de acordo com a Resolução CoPGr 6018 de 2011

Orientador:

Prof. Dr. **PETERSON RICARDO FIORIO**

Dissertação apresentada para obtenção do título de  
Mestre em Ciências. Área de concentração: Engenharia  
de Sistemas Agrícolas

Piracicaba  
2022

**Dados Internacionais de Catalogação na Publicação**  
**DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP**

Ribeiro, Luiz Augusto Leite

Utilização de dados hiperespectrais para diferenciação de variedades de cana-de-açúcar (*Saccharum officinarum* L.) / Luiz Augusto Leite Ribeiro. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2022.

38 p.

Dissertação (Mestrado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Sensoriamento remoto 2. Sensoriamento proximal 3. Random forest 4. Análise de Componentes Principais (PCA) I. Título

## DEDICATÓRIA

Dedico este trabalho ao meu filho, à minha  
companheira, à minha mãe e ao meu falecido pai.

## AGRADECIMENTOS

Agradeço a todos que me acolheram nestes anos intensos e marcantes.

Cinco dias antes do início das atividades de mestrado a perda de meu pai, José Roberto Leite Ribeiro, agradeço do fundo do meu coração e com muita saudade o seu apoio.

Ao meu falecido sogro que antes de partir me acolheu em sua casa no início dessa nova jornada.

À minha família Guido, meu filho, Ângela amor da minha vida, a minha mãe Regina, a minha sogra Dona Beatriz, e aos meus irmãos Marcelo, Zé, minha cunhada Ana, aos sobrinhos Lucas e Stella por todo suporte.

Ao Professor Dr. Peterson Ricardo Fiorio por sua dedicação e confiança e Dr. Rodnei Rizzo pela imensa ajuda nesta caminhada.

À Agência Paulista de Tecnologia dos Agronegócios (APTA), em especial ao Dr. Fábio Luís Ferreira Dias por toda generosidade em transmitir conhecimento, auxílio e por permitir que o experimento fosse realizado.

À Professora Dr. Érica Nakai pelas palavras amigas nas horas difíceis, pelos ensinamentos e sempre estar lá para me reconduzir rumo ao norte.

Ao Ricardo Michelotto, aquele cara durão, mas que no fundo é uma manteiga, pelas conversas e palavras amigas.

À minha colega e amiga Michaela Barbara Neto por todo conhecimento compartilhado e ajuda nos trabalhos de campo.

Ao Topogeo, Natália, Barbara, Larissa, Arthur pela imensa ajuda nos trabalhos de campo.

Aos meus amigos de vida Maestro Otávio Simões, Tiago Nicolosi, João Henrique Talocchi, Ricardo Pastore, Zanta, Louis, Guti, Goma, Kitos, Dutcho, Nebu, Leo, Tonhão, Nou, Piro, Piranha, Daceda, seu Petreli.

Ao Bruno Balboni, vulgo carioca, grande fazendeiro de vasilhões.

Aos meus colegas e chefes da Fairfax Seguros.

Aos professores Dr. Leandro Gimenez, Dr. Thiago Romanelli pelas orientações e conversas ao longo desta etapa de minha vida.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

À FAPESP pela aquisição do equipamento FieldSpec® 3 Jr através do projeto de pesquisa no 2013/22435-9.

## EPÍGRAFE

*“No piquete relincha um poldro; um galo álaçre  
tatala a asa triunfal, ergue a crista de lacre,  
clarina a recolher entre varas de cerdos  
mexem-se ruivos bois processionais e lerdos  
e num magote escuro a manada se abisma  
na treva.*

*Anoiteceu.*

*Juca Mulato cisma.”.*

*Menotti Del Picchia*

## SUMÁRIO

RESUMO .....	7
ABSTRACT .....	8
1. INTRODUÇÃO .....	9
Referências .....	12
2. UTILIZAÇÃO DE DADOS HIPERESPECTRAIS PARA DIFERENCIAÇÃO DE VARIEDADES DE CANA-DE-AÇÚCAR ( <i>Saccharum officinarum</i> L.) .....	15
2.1. Introdução .....	15
2.2. Material e Métodos .....	16
2.2.1. Localização da área de estudo .....	16
2.2.2. Descrição do experimento .....	17
2.2.3. Dados meteorológicos .....	18
2.2.4. Descrição da amostragem .....	19
2.2.5. Dados Espectrais da Folha .....	20
2.2.6. Pré-processamento .....	20
2.2.7. Análise de componentes principais (PCA) .....	21
2.2.8. Seleção das Variedades .....	21
2.2.9. Classificação das Variedades .....	22
2.3. Resultados e discussão .....	22
2.3.1. Seleção das Variedades .....	22
2.3.2. Análise das Curvas Médias brutas e da primeira derivativa .....	23
2.3.3. Análise das Componentes Principais (PCA) .....	28
2.3.4. Classificação das variedades de Cana-de-Açúcar .....	32
2.4. Conclusões .....	33
Referências .....	34

## RESUMO

### **Utilização de dados hiperespectrais para diferenciação de variedades de cana-de-açúcar (*Saccharum officinarum* L.)**

A cana-de-açúcar remete aos primórdios da colonização do Brasil, sendo uma cultura na qual o Brasil possui destaque mundial e é o maior produtor. Para o país se manter nesta posição tornam-se necessários contínuos estudos na área de melhoramento genético. A identificação e mapeamento constantes destas áreas é fundamental para o pagamento dos *royalties* pelo uso das variedades melhoradas. Técnicas rápidas, com alta taxa amostral e que permitam monitoramento de cana-de-açúcar tornam-se necessárias para estimar a área plantada de cada variedade. Estudos com sensores hiperespectrais e aprendizado de máquinas vêm sendo empregados em diversas áreas da agricultura, inclusive em estudos de cana, com bons resultados, podendo ser promissoras para a discriminação de variedades. Desta maneira este trabalho teve como objetivo discriminar variedades de cana-de-açúcar utilizando dados hiperespectrais e técnicas de aprendizado de máquina e Random Forest junto com a análise de componentes principais (PCA). As variedades selecionadas para este estudo foram a IACSP 95-5094, IACSP 01-3127 e IACSP 96-2042, sendo que as regiões que tiveram a maiores influências no agrupamento dos materiais foram a região ligada ao red-edge (725 nm), muito correlacionada ao pigmento clorofila e nitrogênio foliar e as regiões de 1400 nm e 1900 nm correlacionadas com o conteúdo de água nas folhas, sendo um bom indicativo da adaptabilidade ao ambiente de produção. Utilizando apenas dados hiperespectrais do visível e do red-edge (450 a 800 nm) no algoritmo Random Forest, o presente estudo alcançou a acurácia de 71,72% e índice Kappa de 0,57, indicando bom potencial na utilização de dados hiperespectrais na classificação de variedades de cana-de-açúcar.

Palavras-chave: Sensoriamento proximal, Random forest, Análise de Componentes Principais (PCA)



## ABSTRACT

**Use of hyperspectral data to differentiate sugarcane varieties (*Saccharum officinarum* L.)**

Sugarcane refers to the beginnings of the colonization of Brazil, being a culture in which Brazil has a worldwide prominence and is the largest producer. For country or country, keeping in this position makes the studies continuous and the best genetic area. The identification and mapping of these areas is essential for the payment of royalties for the use of improved varieties. Fast sampling techniques, in order to make the high pick rate possible to estimate the area of each variety. Studies with hyperspectral sensors and machine learning have been used in several areas of agriculture, including sugarcane, with good results, which may be promising for a variety of varieties. This way of working aimed to discriminate sugarcane varieties using hyperspectral data and Random Forest machine learning techniques together with a principal component analysis (PCA). The varieties selected for this study were made in IACSP 95-5094, IACSP 01-3127 and IACSP 96-2042, and the regions that had the influence on the grouping of materials were in the region related to the red-edge (725 nm), very The chlorophyll pigment is correlated to the leaf pigment and to the regions of 1400 nm and 1900 nm correlated with the water content in the leaves, being a good indication of the adaptability to the production environment. Hiperspect Use only visible and red-edge (450 to 800 nm) without Random Forest data, the present study reached an accuracy of 71.72% and Kappa index of 0.57, indicating good potential in the use of hyperspectral data in classification of sugarcane varieties.

Keywords: Proximal remote sensing, Random forest, Principal component analysis (PCA).

## 1. INTRODUÇÃO

A cana-de-açúcar tem grande importância econômica para o Brasil, sendo atualmente o maior produtor do mundo (DOE, 2021), e segundo na produção de álcool (USDA, 2021). O país possui cerca de 10 milhões de hectares plantados que produziram 654 milhões de toneladas de cana na safra 2020/21 (IEA, 2022), com produtividade média entre 2010 e 2018 de 75 toneladas por hectare (FAO, 2021), apesar de atualmente ser o líder na produção desta cultura, esta cultura ao longo de sua história no país passou por inúmeros momentos de crise, sendo em boa parte em consequência das baixas produtividades que ocasionaram baixa competitividade entre os países produtores, resultado do uso variedades não adaptadas as condições edafoclimáticas, suscetíveis a pragas e doenças (RODRIGUES; ROSS, 2020).

Apesar deste cenário passado o Brasil volta a liderar a produção de cana no século XX, fato devido a estruturação do setor, e como um dos pilares de sua reorganização o desenvolvimento e fomento de programas de melhoramento genético para a cultura, permitindo a criação de variedades resistentes a pragas e doenças, adaptadas as condições edafoclimáticas das regiões tradicionais e de expansão da cultura (RODRIGUES; ROSS, 2020), motivo pelo excelente desempenho do setor sucroalcooleiro brasileiro em relação aos demais países produtores, porém nos dias atuais verifica-se diminuição do financiamento público nos programas, e os recursos necessários para a manutenção para os programas advindos do pagamento de royalties pelo uso de suas variedades (FRANCISCO, 2014).

Visto a quantidade de áreas de produção e quantidade de produtores envolvidos é fundamental o monitoramento e mapeamento com cada variedade de cana utilizada, a fim de garantir recursos para os programas de melhoramento. Assim, a utilização de técnicas de sensoriamento remoto vem a contribuir agilizando o processo com a possibilidade de mapear vastas extensões de maneira rápida e precisa, reduzir custos, por não necessitar de pessoas treinadas para identificação das variedades, ou mesmo a utilização de análises genéticas e bioquímicas caras e demoradas (ALMEIDA; ROCHELLE; CROCOMO, 1995).

Os sensores remotos conseguem detectar as interações do espectro eletromagnético com a vegetação. Cada comprimento de onda da energia eletromagnética do sol interage de forma específica com a superfície foliar, refletindo em uma assinatura espectral específica para cada organismo. Estas peculiaridades se devem às diferenças nas concentrações de pigmentos que atuam na região do visível (400 nm a 700 nm) e que são relacionados ao processo fotossintético e de fotoproteção (MOREIRA, 2011; MISHRA et al. 2017), devem-se também à anatomia das folhas que influenciam principalmente os processos de reflectância no infravermelho próximo (MISHRA et al. 2017; FURLANETTO et. al., 2018) e na região do infravermelho médio (1100 nm a 2500 nm) devido principalmente a presença de água (MISHRA et al. 2017).

Moreira (2011) afirma que estas diferenças componentes e bioquímicos e estruturas dos vegetais refletem em diferenças nas reflectâncias ao longo do espectro de espécie para espécie, sendo necessária a utilização de sensores que consigam detectar estas diferenças e também de técnicas estatísticas multivariadas para auxiliar na interpretação destas informações.

Atualmente os estudos com sensores remotos focam principalmente parâmetros biofísicos, índices de vegetação para a mensuração de biomassa, pigmentos, sanidade das culturas, estresses bióticos e abióticos, permitindo um melhor entendimento dos processos que ocorrem em diferentes espécies vegetais, o que vem sendo aprimorado nas últimas décadas devido a evolução destes sensores possibilitando uma ampliação de suas aplicações na área agrônoma (JENSEN, 2011).

Sensores multiespectrais e hiperespectrais estão mais presentes na agricultura, pois possibilitam a aquisição de informações sobre as condições das áreas agricultáveis constantes e em diversas escalas, melhorando a qualidade da informação auxiliando no planejamento e tomadas de decisão por parte dos produtores ou demais agentes da cadeia de produção (JENSEN, 2011).

Os estudos sobre essa tecnologia englobam várias temáticas, como fertilidade do solo, espécies invasoras, déficit hídrico, identificação de pragas e doenças, além de modelos matemáticos a fim de caracterizar problemas específicos de cada cultura (FURLANETTO et. al., 2018; ZHANG e ZHOU, 2015), utilização índices de vegetação para a caracterização de parâmetros biofísicos, como área foliar, fitomassa, radiação fotossinteticamente ativa absorvida, produtividade das culturas plantas (FURLANETTO et. al., 2018) e caracterização de deficiência nutricional em plantas (FURLANETTO et. al., 2018; MARTINS, 2016; BARROS, 2016), conteúdo relativo de água e espessura equivalente de água nas folhas de diferentes espécies vegetais (BONILLA, 2015; ZHANG E ZHOU, 2015; STRABELLI, 2016).

No que se refere à classificação e identificação, o sensoriamento remoto vem sendo principalmente utilizado na distinção entre espécies cultivadas e espécies de plantas daninhas, visto a necessidade de um controle mais preciso e utilização mais racional de insumos agrícolas. Segundo Kumar et al. (2010), apesar das assinaturas espectrais de espécies vegetais serem similares, há diferenças sutis entre elas, mudanças nas posições relativas e diferenças nas profundidades de absorção que permitem, ao se cruzarem essas informações, a discriminação dessas curvas. Henry, Shaw e Bruce (2004) utilizando análise discriminante linear, conseguiram separar espécies invasoras da cultura da soja independentemente do status hídrico da cultura de interesse com as plantas daninhas. O trabalho verifica que com o aumento do estresse hídrico houve melhorias na discriminação entre soja e as plantas daninhas, tendo a melhor discriminação entre soja e *Cassia obtusifolia*. Souza et al. (2019) conseguiram separar plantas daninhas da cultura da cana-de-açúcar obtendo quase 100% de acurácia.

Poucos trabalhos tratam da discriminação de cultivares de culturas agrícolas, sendo que a maioria é utilizada para auxiliar programas de melhoramento genético e estudar as metodologias empregadas, como o trabalho realizado por Zhu et al. (2019) no qual o utilizam imagens hiperespectrais de sementes de soja para separar 3 variedades, aplicando a técnica de rede neural convolucional e análise de componentes principais com intuito de encontrar o número mínimo de amostras para o treinamento do modelo de classificação, que concluiu que com 60 amostras para cada cultivar era possível obter 96,33% de acurácia. Outro trabalho que pode ser mencionado foi realizado por Li et al. (2021), cujo objetivo foi testar e diferentes técnicas de aprendizagem de máquinas, como cluster k-média, pls, vetores de suporte e rede neural convolucional, e verificar quais destas possuía a maior eficiência em classificar de 2 a 4 cultivares de soja. Os autores concluíram que a técnica de rede neural convolucional é a mais estável em relação a acurácia nos diferentes experimentos empregados no trabalho e com a melhor acurácia em todos os casos em relação aos demais algoritmos.

Crussiol et al. (2018) tiveram bons resultados na discriminação de diferente cultivares de soja sob diferentes regimes hídricos, além de possibilitar a verificação de questões de adaptabilidade aos diferentes ambientes de produção por parte de cada material, tendo acurácias para cada variedade variando de 50 a 100%, sendo que as variedades com maior confusão entre si foram com matérias com um *background* genético mais similar. Junior et al. (2018) exploram mais a identificação de cultivares de soja e criação de um banco de dados de assinaturas espectrais a fim de subsidiar pagamentos de royalties que tiveram excelente resultado na discriminação de 4 cultivares com acerto por volta dos 99% de acurácia global (BMX Potência, NA5909, Dom Mario e FT Campo Mourão).

Porém, para cana-de-açúcar foram encontrados poucos estudos sobre discriminação de variedades na literatura, sendo que a maioria utiliza sensores remotos orbitais multiespectrais e hiperespectrais, e alguns estudos com sensores hiperespectrais proximais.

Fortes e Demattê (2007) tentam discriminar variedades de cana-de-açúcar utilizando imagens do sensor ETM do satélite Landsat7 e obtiveram bons resultados em separar 4 variedades, sendo que o melhor resultado foi de 100% foi para a variedade RB855536, e o pior para a cultivar SP81-3250 com 89,19% de acerto. Nesse estudo, além das reflectâncias das bandas da região do azul (B1), do verde (B2), do vermelho (B3), do infravermelho próximo (B4) e infravermelho de ondas curtas (B5), utilizou-se o auxílio da dispersão entre os valores de GNDVI e de reflectância da banda 4, sendo que os índices de vegetação provaram ser ferramentas úteis na discriminação de cultivares.

Um estudo realizado por Galvão, Formaggio e Tisot (2007), visando ver os efeitos da resolução espectral de alguns sensores (AVHRR 3/NOAA 17, CCD/CBERS 2, HRV-IR/SPOT 5, ETM+/LANDSAT 7, MODIS/TERRA, ASTER/TERRA, HYPERION/EO-1) na discriminação das variedades de cana-de-açúcar, mostrou bons resultados, sendo que com o aumento da resolução espectral do sensores houve considerável ganho na acurácia da classificação das variedades de cana-de-açúcar, sendo a pior acurácia obtida para o sensor AVHRR 3 com 58% de acurácia e a melhor pelo sensor HYPERION com 87% de acurácia.

Outros estudos com dados hiperespectrais de satélites mostram bons resultados, como o realizado na Austrália utilizando 40 índices de vegetação extraídos de imagens do satélite Hyperion, que objetivou mapear 5 e 8 variedades de cana-de-açúcar e determinar a separabilidade e acurácia desta classificação utilizando análise discriminante. Os autores relatam que tiveram bons resultados de 97,5% quando utilizaram apenas 5 variedades e 72,4% quando utilizaram as 8 variedades (APAN et al., 2004). Outro estudo, realizado no Brasil com o mesmo satélite, valeu-se 5 variedades de cana-de-açúcar utilizando análise discriminante múltipla (MDA) e índices de vegetação, sendo que umas das cultivares por possuir grande diferença de sua curva espectral para as demais cultivares foi discriminada utilizando um simples limiar na banda do infravermelho próximo. As demais variedades foram discriminadas entre si obtendo uma acurácia de 87,5%, os autores indicam que diferenças na estrutura do dossel são fatores que ajudaram na distinção das cultivares via sensoriamento remoto orbital (GALVÃO, FORMAGGIO, TISOT; 2005).

Já estudos com sensores hiperespectrais proximais visando a separação de variedades de cana-de-açúcar são mais escassos, mas mostram a potencialidade na utilização destes sensores com o intuito de discriminação de variedades de cana-de-açúcar, como o estudo de Johnson et al. (2008) que utilizaram dados hiperspectrais e de concentração de pigmentos de plântulas de cana de 5 variedades comerciais e 2 espécies de cana-de-açúcar para a classificação utilizando análise discriminante. Os autores informam que tiveram resultados de 95 a 100% de classificações corretas e, enquanto utilizando dados dos pigmentos, 76 a 81% de classificação correta.

Outro estudo foi o de Steidle Neto *et al.* (2018) que ao invés de utilizar dados de folhas utilizam dados hiperespectrais do colmo, utilizando a faixa 450 a 1000 nm do colmo de 4 variedades de cana-de-açúcar (RB867515, RB855453, RB928064 e RB92579) e técnicas multivariadas para auxiliar na interpretação dos resultados. Os autores observam que a utilização da análise de componentes principais não conseguiu separar corretamente todas as variedades, havendo grande sobreposição no gráfico de dispersão da componente principal 1 e 2 entre a RB867515 e RB855453. Por outro lado, utilizando análise por mínimos quadrados parciais, análise discriminante fatorial e análise discriminante stepwise 81%, 82% e 76% de classificações corretas, sendo que a utilização, os mesmos autores constatam que as regiões do visível e do infravermelho próximo entre 600 a 750 nm foram as que tiveram maior relevância no estudo.

Visto escassez de estudos com cana-de-açúcar e sua potencialidade na utilização de equipamentos hiperespectrais e técnicas multivariadas para a discriminação de variedades como ferramenta de auxílio ao monitoramento de variedades de cana-de-açúcar, este trabalho teve como objetivo discriminar variedades de cana-de-açúcar utilizando dados hiperespectrais e técnicas de aprendizado de máquinas.

## Referências

- ALMEIDA, M. de; ROCHELLE, L.A.; CROCOMO, O. J. Chave analítica para determinação de dez variedades de cana-de-açúcar (*Saccharum spp.*). **Scientia Agricola**. Piracicaba, v.52, n. 1, p. 16-19, jan./abr., 1995.
- APAN, A; HELD, A.; PHINN, S.; MARKLEY, J. Spectral discrimination and classification of sugarcane varieties using EO-1 Hyperion Hyperspectral Imagery. Proceedings of the 25th Asian Conference on Remote Sensing (ACRS 2004). **Asian Association on Remote Sensing**, 2004, pp. 1–5.
- BARROS, P.P. da S. Dados hiperespectrais de dossel e sua correlação com o nitrogênio aplicado a cultura da cana-de-açúcar. 2016. 131 p. Tese (Doutorado em Engenharia de Sistemas Agrícolas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2016.
- BONILLA, M.M.Z. Dados hiperespectrais na determinação do conteúdo relativo de água na folha em cana-de-açúcar. 2015. 86 p. Dissertação (Mestrado em Engenharia de Sistemas Agrícolas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2015.
- CRUSSIOL, L.G.T.; NANNI, M.R.; FURLANETTO, R.H.; CEZAR, E.; SIBADELLI, R.N.R.; MERTZ-HENNING, L.M.; NEPOMUCENO, A.L.; NEUMAIER, N.; FARIAS, J.R.B. Assinatura espectral de cultivares de soja submetidas a diferentes níveis de disponibilidade hídrica. In: VIII Congresso Brasileiro de Soja, 2018, Goiânia. VIII Congresso Brasileiro de Soja, Goiânia, GO - 2018 : inovação, tecnologias digitais e sustentabilidade da soja: anais. Brasília: Embrapa, 2018. v. 1. p. 293-295.
- FAO - Food and Agriculture Organization. Statistics Division. Area harvested, yield and sugarcane production in the world. Disponível em: <<http://www.fao.org/faostat/en/#data/QC>>. Acesso em: 15 jan. 2021.
- FORTES, C.; DEMATTÊ, J.A.M. Discrimination of sugarcane varieties using Landsat 7 ETM+ spectral data. **International Journal of Remote Sensing**. V. 27, n. 7, p.1395-1412, abr., 2007.
- FRANCISCO, A. C. O regime da propriedade intelectual para a Cana-de-Açúcar. Tese. Faculdade de Direito, Universidade de São Paulo, São Paulo. 2014.
- FURLANETTO, R. H. ; NANNI, M. R. ; CRUSSIOL, L. G. T. ; OLIVEIRA JUNIOR, A. . Índices de vegetação obtido por sensor hiperespectral terrestre na identificação da deficiência de potássio na cultura da soja. In: VIII Congresso Brasileiro de Soja, 2018, Goiania. VIII Congresso Brasileiro de Soja, Goiânia, GO - 2018 : inovação, tecnologias digitais e sustentabilidade da soja: anais. Brasília: Embrapa, 2018. v. 1. p. 820-822.
- GALVÃO, L.S.; FORMAGGIO, A.R.; TISOT, D.A. Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data. **Remote Sensing of Environment**, v.94, n.4, p.523–534, 2005. Disponível em:<<https://doi.org/10.1016/j.rse.2004.11.012>>.
- GALVÃO, L.S.; FORMAGGIO, A.R.; TISOT, D.A. The influence of spectral resolution on discriminating Brazilian sugarcane varieties. **International Journal of Remote Sensing**. v.27, n.4, p. 769-777 2007. Disponível em:<[10.1080/01431160500166011](https://doi.org/10.1080/01431160500166011)>.
- HENRY, W.B.; SHAW, D.R.; BRUCE, L.M. Spectral reflectance curves to distinguish soybean from common cocklebur (*Xanthium strumarium*) and sicklepod (*Cassia obtusifolia*) grown with varying soil moisture. **Weed Science**, v. 52, p.788-796, set.-out. , 2004.

- IEA-Instituto de economia agrícola. Alta na Produção e nas Exportações de Açúcar Marca a Safra 2020/21 de Cana. Disponível em:<<http://www.iewa.gov.br/out/TerTexto.php?codTexto=15925>>. Acesso em: 30 jan. 2022.
- JENSEN, J.R. Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestre. 3. ed. São José dos Campos: Parêntese, 2011. 598 p.
- JOHNSON, R.M.; VIATOR, R.P.; VEREMIS, J.C.; RICHARD, E.P.; ZIMBA, P. V. Johnson et al: Discrimination of sugarcane varieties with pigment profiles and high resolution, hyperspectral leaf reflectance data. p.63–75, [s.d.].
- JUNIOR, C.A.S.; NANNI, M.R.; SHAKIR, M.; TEODORO, P.E.; OLIVEIRA-JUNIOR, J.F.; CEZAR, E.; GOIS, G.; LIMA, M.; WOJCIECHOWSKI, J.C.; SHIRATSUCHI. Soybean varieties discrimination using non-imaging hyperspectral sensor. **Infrared Physics & Technology**. V. 89, p.338-350, fev, 2018. Disponível em <<https://doi.org/10.1016/j.infrared.2018.01.027>>
- KUMAR, L.; SKIDMORE, A. K.; MUTANGA, O. Leaf level experiments to discriminate between eucalyptus species using high spectral resolution reflectance data: use of derivatives, ratios and vegetation indices. **Geocarto International**. V. 25, n. 4, p. 327-344, 2010. Disponível em <<https://doi.org/10.1080/10106040903505996>>.
- LI, H.; ZHANG, L.; SUN, H. Identification of soybean varieties based on hyperspectral imaging technology and one-dimensional convolutional neural network. p.1–14, 2021. Disponível em <<https://doi.org/10.1111/jfpe.13767>>.
- MARTINS, J.A. Dados hiperespectrais para predição do teor foliar de nitrogênio em cana-de-açúcar.2016. 65 p. Tese (Doutorado em Engenharia de Sistemas Agrícolas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2016.
- MISHRA, P. ASAARI, M.S.M.; HERRERO-LANGREO, A.; LOHUMI, S.; DIEZMA, B.; SCHEUNDERS. Close range hyperspectral imaging of plants: A review. **Biosystems Engineering**. V. 164, p. 49-67, dez., 2017. Disponível em: <<https://doi.org/10.1016/j.biosystemseng.2017.09.009>>.
- MOREIRA, A. M. Fundamentos do sensoriamento remoto e metodologias de aplicação. 4. ed. Viçosa: UFV, 2011. 422 p.
- FELIPE, M.; SOUZA, D.; RIOS, L.; ROBSON, S.; OLIVEIRA, D.M.; ANTONIO, M.; COUTINHO, N.; FERREIRA, C. ScienceDirect Spectral differentiation of sugarcane from weeds. **Biosystems Engineering** v.0, p.2–7, 2019. Disponível em: <<https://doi.org/10.1016/j.biosystemseng.2019.11.023>>.
- STEIDLE NETO, A.J.; LOPES, D.C.; TOLEDO, J. V.; ZOLNIER, S.; SILVA, T.G.F. Classification of sugarcane varieties using visible/near infrared spectral reflectance of stalks and multivariate methods. **Journal of Agricultural Science**, v.156, p.537–546, 2018. Disponível em:<<https://doi.org/10.1017/S0021859618000539>>.
- STRABELI, T.F. Resposta hiperespectral na determinação do conteúdo de água na folha em diferentes espécies de Eucalyptus spp.2016. 108 p. Dissertação (Mestrado em Engenharia de Sistemas Agrícolas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2016.
- USDA. Foreign Agricultural Service. Sugar: World Markets and Trade, November, 2020. Disponível em <<https://apps.fas.usda.gov/psdonline/circulars/Sugar.pdf>>. Acesso em: 15 jan. 2021.
- ZHANG, F.; ZHOU, G. Estimation of Canopy Water Content by Means of Hyperspectral Indices Based on Drought Stress Gradient Experiments of Maize in the North Plain China. **Remote Sensing**. V, 7, p. 15203-15223, nov., 2015. Disponível em <<https://doi.org/10.3390/rs71115203>>

ZHU, S; ZHOU, L; ZHANG, C; BAO, Y; WU, B; CHU, H; YU, Y; HE, Y; FENG, L. Identification of Soybean Varieties Using Hyperspectral Imaging Coupled with Convolutional Neural Network. **Sensors (Basel)**. 2019 Sep, v. 19, n. 19, p. 4065. Disponível em: <<https://doi.org/10.3390/s19194065>>.

## 2. UTILIZAÇÃO DE DADOS HIPERESPECTRAIS PARA DIFERENCIAÇÃO DE VARIEDADES DE CANA-DE-AÇÚCAR (*Saccharum officinarum* L.)

### 2.1. Introdução

A cana-de-açúcar remete aos primórdios da colonização do Brasil onde apenas duas variedades foram utilizadas durante 370 anos. Durante esse período a produção fora ineficiente, consequência da susceptibilidade a pragas, doenças, exaurimento do solo e baixa tecnologia para obtenção de açúcar (RODRIGUES; ROSS, 2020). Ainda segundo os autores, devido ao aumento da importância da cultura, no século XIX surgem os primeiros estudos para o melhoramento vegetal da cana-de-açúcar, que impulsionaram no século seguinte a estruturação do setor de produção, com criação de programas de melhoramento vegetal, assim desenvolvendo inúmeras variedades adaptadas as diversas condições edafoclimáticas do Brasil.

Hoje o Brasil é o maior produtor de cana-de açúcar do mundo (DOE, 2021), e segundo na produção de álcool (USDA, 2021). O país possui cerca de 10 milhões de hectares plantados que produzem 654 milhões de toneladas de cana na safra 2020/21 (IEA, 2022), com produtividade média entre 2010 e 2018 de 75 toneladas por hectare (FAO, 2021).

Os programas de melhoramento genético proporcionam a manutenção do desempenho do setor produtivo sucroalcooleiro brasileiro, sendo mantidos principalmente por meio do pagamento de *royalties* pelo uso de suas variedades (FRANCISCO, 2014), necessitando de monitoramento e mapeamento das áreas plantadas, a fim de garantir recursos para o contínuo desenvolvimento de pesquisas nessa área.

No entanto, a identificação das múltiplas variedades mostra-se difícil, uma vez que elas são híbridas interespecies do gênero *Saccharum*, tendo como seus principais progenitores seis espécies: *S. officinarum*, *S. robustum*, *S. sinense*, *S. barberi*, *S. spontaneum* e *S. edule* (CHEAVEGATTI-GIANOTTO et al., 2011) e possuem poucas diferenças morfológicas entre si (ALMEIDA; CROCOMO, 1994). Sendo assim, faz-se necessário pessoas bem treinadas para identificação das variedades, ou mesmo a utilização de análises genéticas e bioquímicas caras e demoradas (ALMEIDA; ROCHELLE; CROCOMO, 1995).

Por outro lado, existem as técnicas espectroscópicas (com sensores hiperespectrais), que podem vir a contribuir na diferenciação e identificação de variedades, possibilitando a aquisição de enorme quantidade de informações sem necessidade de preparo do material, de forma rápida e não onerosa (STEDILE NETO et al., 2018).

Análises espectroscópicas vem sendo conduzidas em cana-de-açúcar com diversos objetivos, como quantificar o nitrogênio foliar (MARTINS et al., 2021, BARROS et al., 2022), determinar o conteúdo de água em folhas em diferentes regimes de irrigação (FIORIO et al., 2018; CHEN et al., 2012), conteúdo de água nos colmos e mapeamento do Brix em canaviais (MARAPHUM et al., 2020), detecção de doenças (GRISHAM; JOHNSON; ZIMBA, 2010), fenotipagem (GONÇALVES et al., 2021) e inclusive na discriminação de variedades utilizando a resposta espectral de colmos com sensores proximais (STEIDLE NETO et al., 2018) e sensores orbitais (GALVÃO; FORMAGGIO; TISOT, 2005).

Ressalta-se que a simples extração da informação dos sensores hiperespectrais ou a utilização de índices espectrais de vegetação, não permitem de forma simples a distinção de variedades de cana, não gerando resultados satisfatórios (FRASSON et al., 2007; GALVÃO; FORMAGGIO; TISOT, 2005). Nestes casos é necessário o uso de técnicas de aprendizagem de máquinas, as quais têm grande vantagem ao analisar dados com alta dimensionalidade, não lineares, provendo assim uma melhor acurácia (PARK et al., 2014; LIU et al., 2020). Tais técnicas vem sendo



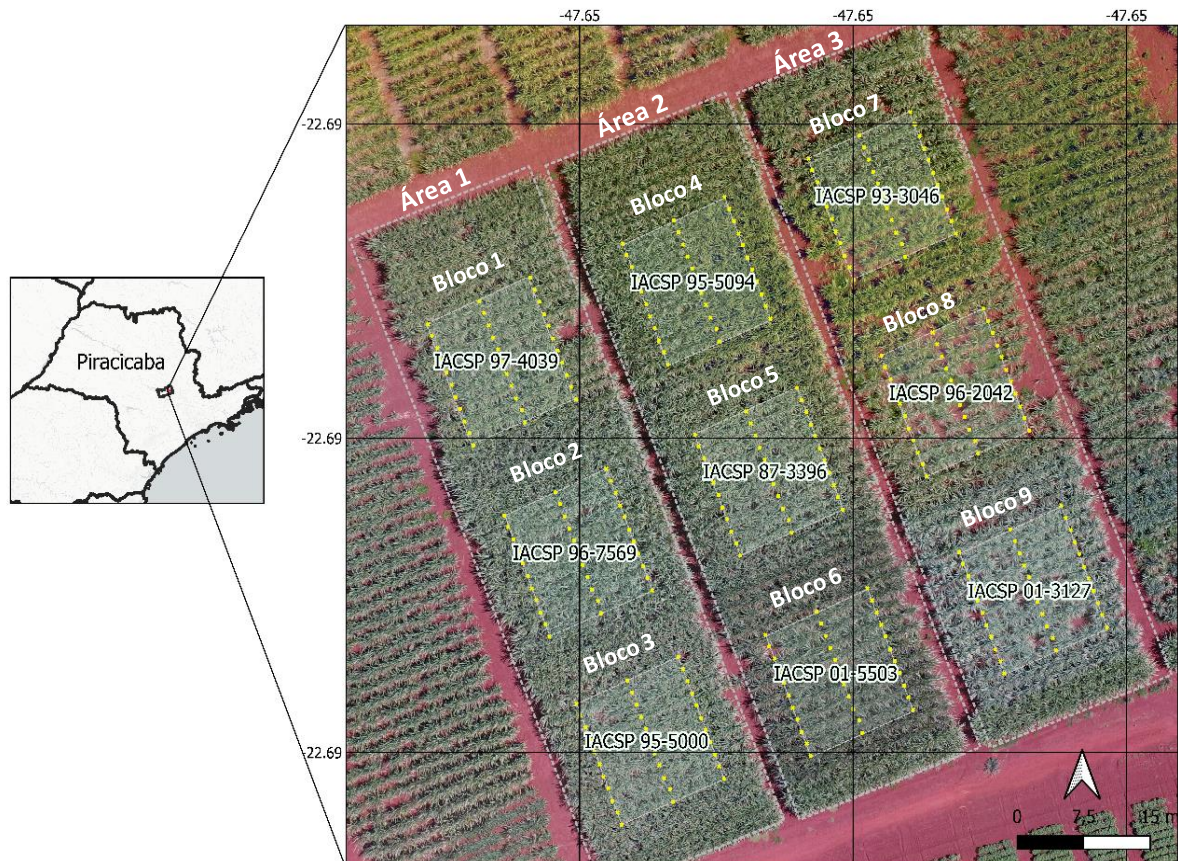
empregadas em diversas áreas da agricultura, como o monitoramento da qualidade da água (LIU et al., 2021; ZHANG et al., 2021), aprimoramento na previsão climática, monitoramento de eventos extremos (DAO; HE; PROCTOR, 2021), apoio ao controle de insetos pragas (SHAO et al., 2021), doenças (BAO et al., 2021; ZHAO et al., 2020), plantas daninhas (SOUZA et al., 2020), previsão e estimativa de safras (ESPINOSA; VESLASQUEZ; HERNANDEZ, 2020; ELAVARASAN et al., 2018), detecção da maturação de frutos (JIE; ZHOU; WEI, 2019), detecção da origem de produtos (ACEVEDO et al., 2007; BONNA et al., 2019), e utilização de inspeção de qualidade de sementes (FABIYI et al., 2019).

Cria-se assim uma expectativa da utilização de sensores hiperespectrais e aprendizagem de máquinas para se tentar diferenciar variedades de cana-de-açúcar no campo de uma forma mais rápida e menos onerosa, com alta densidade amostral. Dentre as técnicas, o aprendizado de máquinas não paramétricas, o Random Forest (RF), vem apresentando bons resultados, com poucos parâmetros a serem ajustados, menor tempo de processamento, baixa superestimativa ou subestimativa de erros e pequena incerteza. Porém esse modelo possui mais problemas de *overfitting* que as demais técnicas (LIU et al., 2020). Outra questão importante é com relação ao número de amostras que, quando reduzidas em comparação ao número de variáveis, podem acarretar performance ruim do modelo, o que eventualmente é resolvido com técnicas de redução de dimensionalidade, como a análise de componentes principais (PCA). Logo, a combinação entre dados hiperespectrais e métodos estatísticos avançados, podem vir a contribuir com melhores performances na classificação de variedades (FABIYI et al., 2019; ZHAO et al., 2020). Sendo assim, este trabalho teve como objetivo discriminar variedades de cana-de-açúcar utilizando dados hiperespectrais e técnicas de aprendizado de máquinas.

## **2.2. Material e Métodos**

### **2.2.1. Localização da área de estudo**

Este estudo foi conduzido na área experimental da Agência Paulista de Tecnologia dos Agronegócios (APTA) de Piracicaba, cujo retângulo envolvente possui as seguintes coordenadas de Latitude 22°41'12" e 22°41'10" Sul e Longitude 47°38'42" e 47°38'45" Oeste (Figura 1).



**Figura 1.** Croqui da área de estudo, sendo constituído de 9 blocos, com 1 variedade por bloco, Imagem do ortomosaico obtida por RPA, com sensor RGB na data (15/01/2019).

### 2.2.2. Descrição do experimento

O solo presente nesta área é um Latossolo Vermelho (LV), o clima da região é denominado subtropical úmido (Cwa), com pluviosidade média anual inferior a 1400 mm, sendo o verão chuvoso e inverno seco, segundo a classificação de Köppen.

A Tabela 1 informa as características de cada variedade de cana-de-açúcar utilizada no experimento inicialmente.

**Tabela 1.** Características das variedades de Cana-de-Açúcar do presente estudo

	<b>IACSP 97- 4039</b>	<b>IACSP 96- 7569</b>	<b>IACSP 95-5000</b>	<b>IACSP 95- 5094</b>	<b>IACSP 87- 3396</b>	<b>IACSP 01- 5503</b>	<b>IACSP 93- 3046</b>	<b>IACSP 96- 2042</b>	<b>IACSP 01- 3127</b>
<b>Colheita</b>	Abr- Ago	Mai-Set	Jun-Out	Jun-Set	Jun-Set	Jun- Out	Jun- Ago	Jul- Nov	Jun- Ago
<b>Ambiente de Produção</b>	B2-E1	B1-D2	A1-C2	A2-C2	C1-E1	B2-D1	A2-C2	A1-C2	A1-C1
<b>Produtividade</b>	Alta	Alta	Muito Alta	Muito Alta	Média	Muito Alta	Alta	Alta	Alta
<b>Brotação de Soqueira</b>	Ótima	Ótima	Ótima	Ótima	Boa	Ótima	Ótima	Boa	Ótima
<b>Perfilhamento</b>	Ótimo	Médio	Ótimo	Ótimo	Médio	Ótimo	Ótimo	Médio	Ótimo
<b>Fechamento de entrelinhas</b>	Ótimo	Bom	Ótimo	Ótimo	Bom	Bom	Ótimo	Bom	Ótimo
<b>Tombamento</b>	Médio	Médio/Alto	Não	Não	Médio/Elevado	Não	Não	Não	Não
<b>Teor de Sacarose</b>	Muito Alto	Alto	Alto	Alto	Médio	Bom	Muito Alto	Alto	Muito Alto
<b>Teor Fibra%/Cana</b>	Médio	Médio	Médio	Médio/Alto	Médio	Médio	Médio	Baixo	Alto

Fonte: Adaptado de Programa Cana (2012); Programa Cana (2016).

O experimento foi implementado em abril de 2018 (cana planta), sendo que os dados utilizados nesse trabalho foram obtidos entre os meses de novembro (14/11/2018) e dezembro (11/12/2018). A correção do solo utilizou-se de calcário, enquanto que a adubação foi realizada com 500 kg/ha do formulado NPK 08-28-16. Já o controle de plantas daninhas foi conduzido com a aplicação de Sulfentazona de 300 g/ha em pré-emergência anteriormente ao plantio, assim como quinze dias após o plantio em pós-emergência.

### 2.2.3. Dados meteorológicos

Conforme os dados da Estação Meteorológica da Escola Superior de Agricultura Luiz de Queiroz (ESALQ), a precipitação acumulada de abril até a primeira coleta foi de 449 mm, com acumulado de 79,2 mm do dia 14/10/2018 até dia 13/11/2018, 1 dia antes da primeira coleta, sendo que ocorreram 10 dias sem precipitação na área de estudo, já da primeira coleta até a segunda coleta houve o acumulado de 180 mm de chuva (Figura 2).

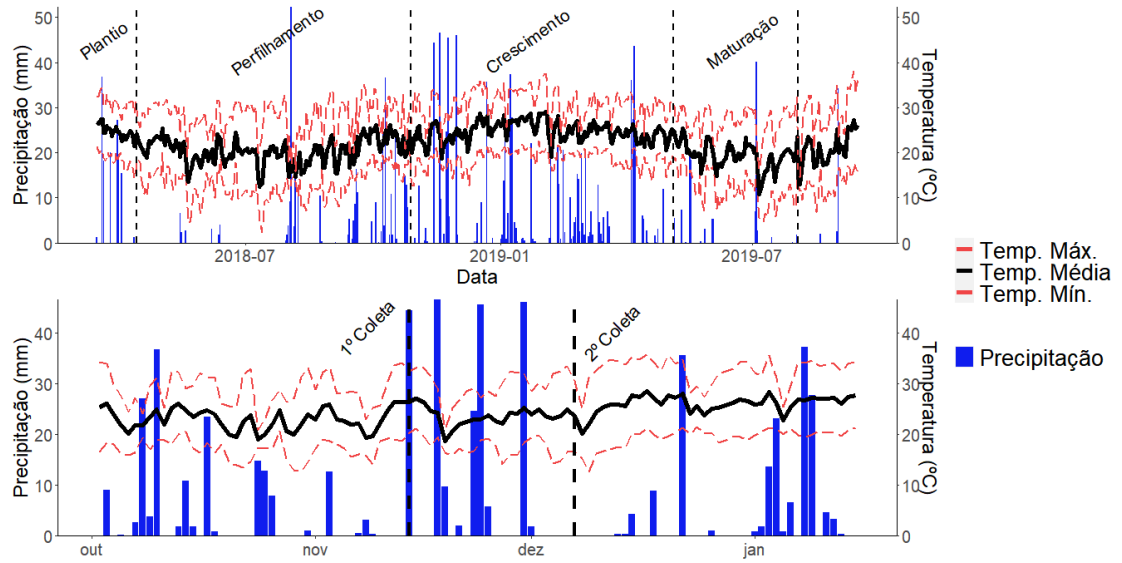


Figura 2. Dados de precipitação e temperatura ao longo do ciclo da cultura.

#### 2.2.4. Descrição da amostragem

A área do experimento é constituída de 3 áreas adjacentes entre si, divididas por carregadores, cada uma com 3 blocos, resultando num total de 9 blocos (Figura 1), sem repetições. Cada área possui 20 metros de largura por 72 metros de comprimento, é constituída por 2 linhas em cada extremo da variedade IACSP 91-1099 e 15 linhas por bloco, totalizando 49 linhas espaçadas em 1,5 metros entre si.

Para cada bloco, foram descartadas 2 linhas de seus extremos, restando 11 linhas centrais. Também foram descartados 5 metros de cada lado das linhas a fim de evitar o efeito de bordadura. Desta forma, as áreas para a coleta de folhas foram constituídas de 15 metros de comprimento por 10 metros de largura.

As cultivares plantadas em cada bloco são: bloco 1 (IACSP 97-4039), bloco 2 (IACSP 96-7569), bloco 3 (IACSP 95-5000), bloco 4 (IACSP 95-5094), bloco 5 (IACSP 87-3396), bloco 6 (IACSP 01-5503), bloco 7 (IACSP 93-3046), bloco 8 (IACSP 96-2042) e bloco 9 (IACSP 01-3127).

Foram realizadas 2 amostragens, correspondendo aos dias 14/11/2018 e 11/12/2018. Nelas foram coletadas, em cada uma dessas datas, 33 folhas por variedade (Bloco) de maneira aleatória, equivalente a 297 folhas por coleta e totalizando 594 folhas.

Foi coletada a primeira folha completamente expandida de cada planta a partir do ápice, denominada como “folha +1” conforme sistema Kuyper (MARAFON, 2012). As folhas foram coletadas, cortadas aproveitando o terço mediano e colocadas em sacos plásticos catalogados, estes acondicionados em caixa térmica com gelo para a manutenção da umidade e conservação dos tecidos foliares (TAVARES et al. 2020). Após as coletas, estas foram levadas ao laboratório e efetuadas as medições de reflectância no espectrômetro (BARROS et al., 2022; MARTINS et al. 2021).

### 2.2.5. Dados Espectrais da Folha

Em laboratório realizou-se a coleta dos dados espectrais de cada folha, evitando-se a coleta sob a nervura central. Para cada folha foram realizadas 10 leituras para obtenção das curvas espectrais médias.

Para a coleta das assinaturas espectrais utilizou-se o aparelho FieldSpec 3 Spectroradiometer (ASD – Analytical Spectral Devices Inc., Boulder, CO, EUA) que opera no intervalo espectral de 350 a 2500 nm, com resolução espectral de 1,4 nm de 350 a 1050 nm e 2 nm de 1050 a 2500 nm (Figura 3 A), acoplado ao Leaf Clip (ASD – Analytical Spectral Devices Inc., Boulder, CO, EUA) para a realização de leituras de reflectância da folha em condições controladas (Figura 3 B), devido ao fato de manter a mesma intensidade de luz e incidência ortogonal (ASD, 2010).

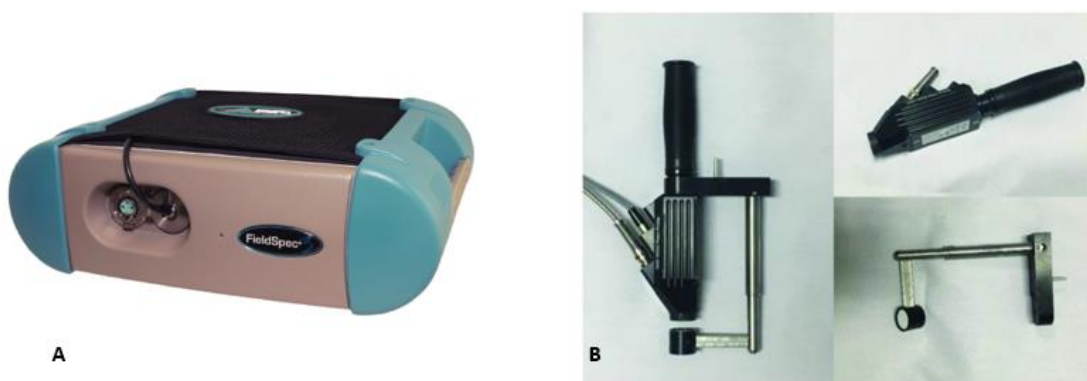
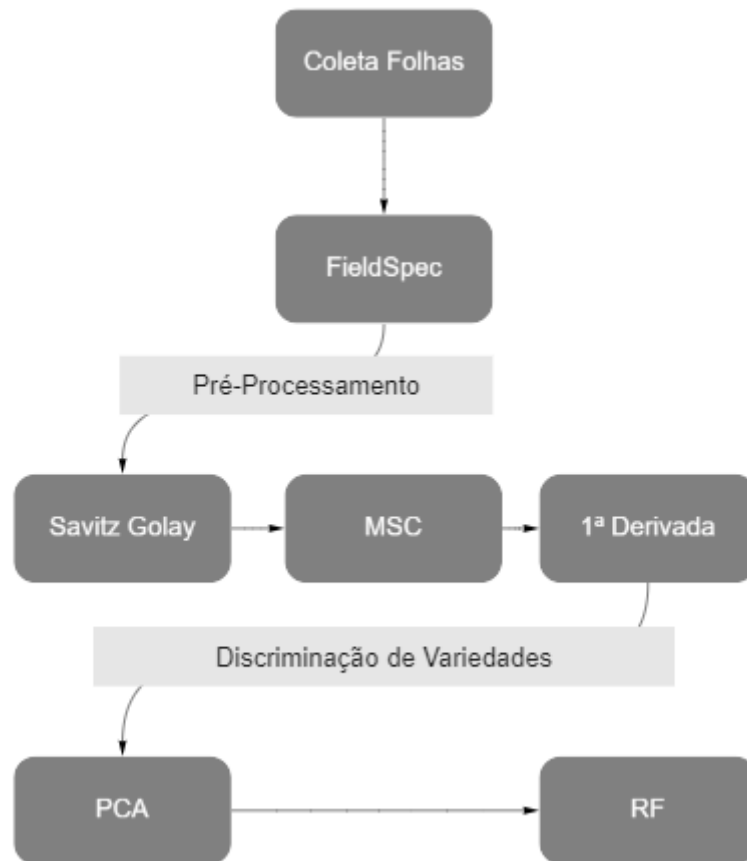


Figura 3. A) FieldSpec Spectroradiometer e B) Leaf Clip.

### 2.2.6. Pré-processamento

Após a obtenção das curvas espectrais estas foram plotadas para a verificação da presença de ruídos e a eliminação destas faixas (ADAMS, 1995; HUA et al.; 2019). A partir desta análise descartou-se os 100 nm iniciais e finais das curvas espectrais, devido a quantidade de ruídos, mantendo para análise a faixa de 450 nm a 2400 nm.

Em sequência a esta etapa, aplicou-se o filtro Savitzky-Golay, do pacote *signal* (LIGGES et al.; 2015) utilizando os parâmetros de polinômio de 1º ordem e janela de 9 pontos (HUANG et al., 2021) para a suavização dos ruídos remanescentes (WEHRENS, 2011). Após a suavização, aplicou o algoritmo MSC (*multiplicative scatter correction*) referente ao pacote *pls* do software R (MEVIK, WEHRENS; 2015). O MSC ajuda a corrigir os efeitos de linha de base (baseline effects) e do espalhamento aditivo e multiplicativo através de uma simples regressão linear (SILALAHY et al.; 2018; GAO et al.; 2021). A partir daí se calculou a primeira derivada com o algoritmo Savitzky-Golay para todas as curvas, com janela de 9 pontos (HUANG et al., 2021).



**Figura 4.** Fluxograma com as etapas de tratamento dos dados.

### 2.2.7. Análise de componentes principais (PCA)

Para visualizar diferenças no comportamento espectral das variedades e a contribuição de cada banda espectral neste processo, utilizou-se a análise de componentes principais, conforme explicado por Wehrens (2011). Trata-se de um método multivariado simples, com o objetivo de encontrar combinações das variáveis capazes de produzir índices e descrever a variação dos dados. Para tanto, transforma-se linearmente um conjunto de dados com variáveis correlacionadas entre si em um conjunto menor de dados com variáveis não correlacionadas (ADANS, 1995; MANLY, 2008; WEHRENS, 2011).

### 2.2.8. Seleção das Variedades

Para seleção das variedades foram utilizados os centroides das componentes principais 1 e 2, que possuísem a maior distância entre si determinada pela plotagem dos centroides no gráfico de dispersão das componentes principais. Para corroborar esta escolha, calculou-se as distâncias euclidianas para todas as cultivares e se escolheu o agrupamento com 3 cultivares que mais se distanciavam entre si.

### 2.2.9. Classificação das Variedades

Para a análise de classificação com o modelo Random Forest optou-se por utilizar apenas as faixas do visível (VIS) e red-edge, devido a maioria dos sensores de campo trabalharem com bandas no VIS-NIR. Outro fator é tentar atenuar os efeitos diretos do conteúdo de água nas folhas, retirando as regiões do infravermelho próximo e principalmente o infravermelho de ondas curtas (em inglês SWIR, ou *Short Wave Infrared*).

O Random forest é um método de agrupamento de dados não paramétrico, baseado na criação de várias árvores de decisão (BREIMAN, 2001). Consiste na utilização de vários classificadores não muito robustos que conjuntamente gerarão melhores resultados do que um único bom classificador. A ideia consiste em criar um enorme número de árvores simples com poucos nós, cada nó criado com um número de variáveis predefinido, utilizando vários subconjuntos de amostras para a criação das arvores, isso permitiu uma maior diversidade dentro do conjunto de dados, com uma melhora nos resultados (WEHRENS, 2011).

Essas árvores são criadas através de várias sub amostras de forma aleatória e com reposição dos dados, técnica denominada amostragem *bootstrap*. Cada árvore de decisão é considerada um classificador, onde serão rotuladas as amostras através de um voto, ou seja, a classe que aparecer mais vezes para cada amostra nas diferentes árvores de decisão será o rótulo desta. As clivagens serão efetuadas com variáveis escolhidas aleatoriamente e através do índice *Gini*, para a escolha do melhor ponto de clivagem (BREIMAN, 2001; ROBNIK-SIKONJA, 2004; CUTLER et al.; 2007).

O pequeno número de amostras e a grande quantidade de variáveis utilizadas, se torna um grande problema para o algoritmo Random Forest, gerando resultados não satisfatórios (LIU et al., 2020; FABIYI et al.; 2019). Para melhorar a performance do modelo, utilizou-se a análise de componentes principais para reduzir a dimensão das variáveis independentes (ZHAO et al., 2020; FABIYI; 2019)

Utilizou-se o pacote Caret do software R para a mensuração da performance do modelo Random Forest (MAX et al.; 2020). Para o parâmetro *ntree* utilizou-se o valor 500 que é o padrão do pacote Caret. Na calibração e treinamento utilizou-se a clivagem entre as coletas para manter a independência das amostras dentro do modelo gerado. No treinamento e calibração do parâmetro *mtry*, utilizou-se como variáveis as 5 primeiras componentes principais, que correspondem a 99,9% da variância total das componentes e a técnica *5 K-fold Cross Validation*. Assim num primeiro momento utilizando a coleta 1 como grupo de treinamento e a coleta 2 como grupo a ser testado na predição, e num segundo momento utilizando a coleta 2 como grupo de treinamento e a coleta 1 como grupo a ser predito.

Para mensuração da qualidade da classificação utilizou o índice Kappa e a Acurácia global.

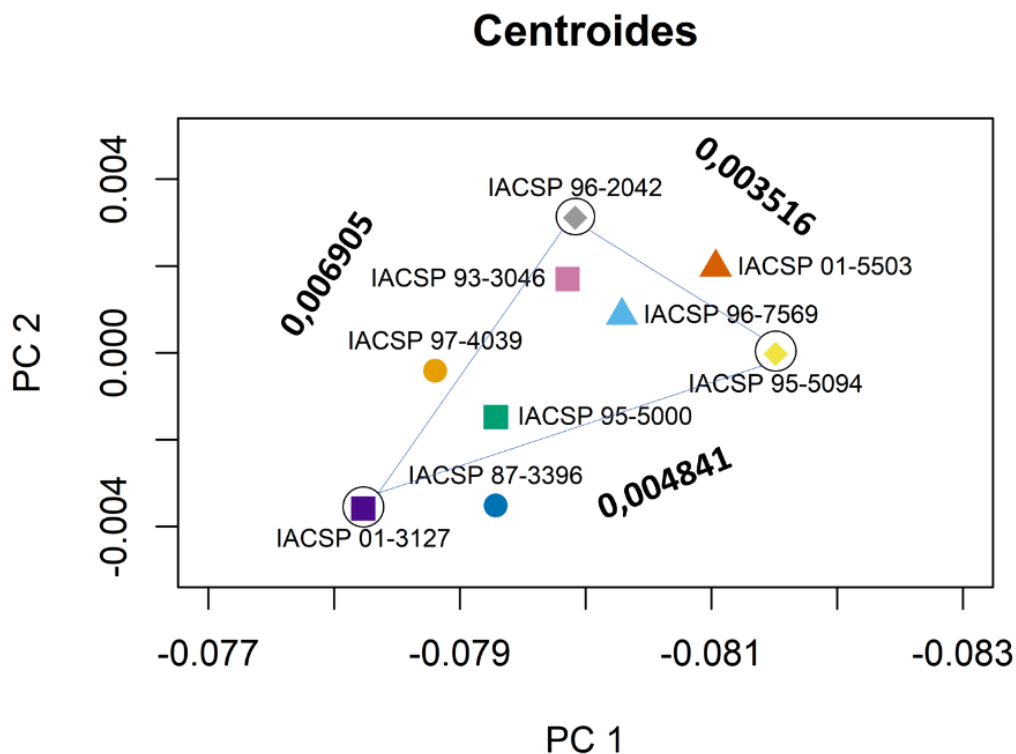
## 2.3. Resultados e Discussão

### 2.3.1. Seleção das Variedades

As variedades deste estudo apresentaram comportamentos espectrais muito parecidos, e no gráfico de dispersão dos scores (Figura 4) todas variedades se sobrepuseram entre si, porém em diferentes graus. Stedile Neto et al. (2018), estudando a resposta espectral do colmo de 4 variedades de cana, verificaram que dois materiais se sobrepunham entre si utilizando a análise de componentes principais, e estes com as piores acurácias do estudo. Os

mesmos autores salientam maiores dificuldades em diferenciar as variedades de cana-de-açúcar utilizando dados espectrais de folhas.

Desta forma, optou-se em selecionar as 3 variedades, dentro dos 9 materiais iniciais, mais distantes entre si, e para isto foram calculados os centroides das componentes principais 1 e 2 e verificadas suas distâncias euclidianas para auxiliar nesta escolha (Figura 4). Como resultado, selecionou-se as canas, IACSP 95-5094 (Bloco 4), IACSP 96-2042 (Bloco 8) e IACSP 01-3127 (Bloco 9) (Figura 1). A variedade IACSP 01-3127 foi a que mais se afastou dos centroides das demais variedades selecionadas, com uma distância de 0,0069 da variedade IACSP 96-2042 e 0,0048 da variedade IACSP 95-5094, enquanto a distância entre os centroides destas duas obteve o valor de 0,0035 (Figura 4).



**Figura 5.** Centroides das componentes principais 1 e 2 para as 9 variedades de cana-de-açúcar.

### 2.3.2. Análise das Curvas Médias brutas e da primeira derivativa

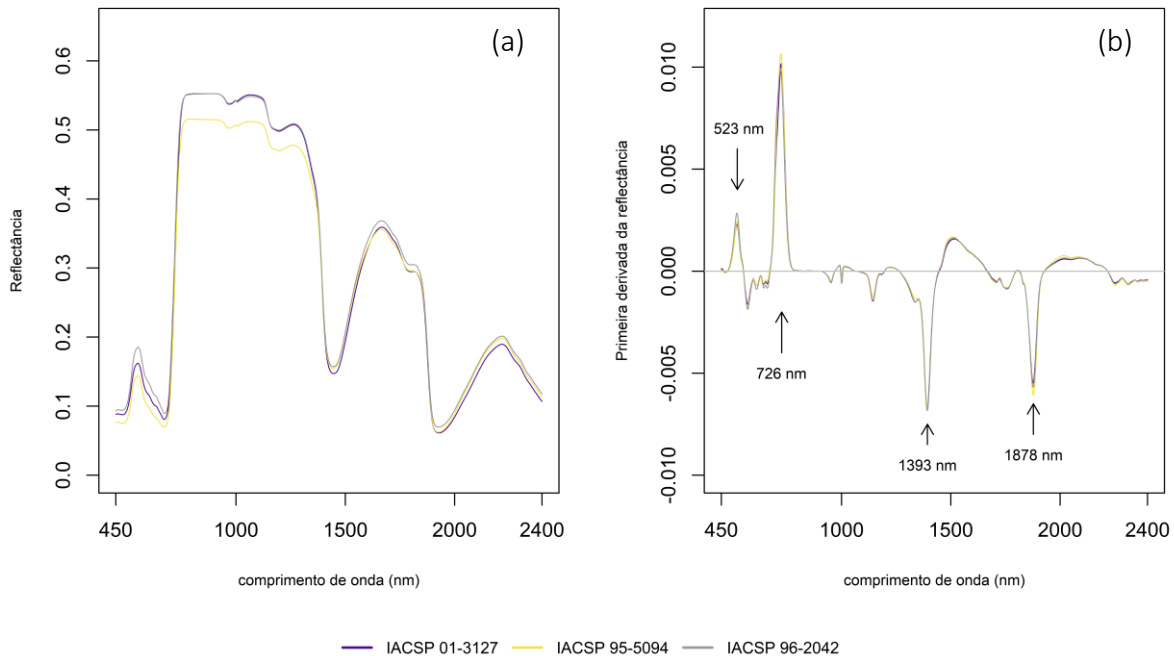
As curvas médias das reflectâncias das coletas 1 e 2 juntas apresentam maiores intensidades para a IACSP 96-2042 ao longo do espectro, seguida da IACSP 01-3127 e da IACSP 95-5094 na região do visível. Nesta região as curvas médias estão visualmente separadas, e no infravermelho próximo (NIR, do inglês *Near Infrared*) as curvas das variedades IACSP 01-2731 praticamente se sobrepõe. No SWIR na faixa dos 1400 nm há uma mudança, a variedade IACSP 95-5094 passa a possuir valores de reflectância intermediários e sua curva média mais próxima da curva da IACSP 01-3127 (Figura 5 A). Já na faixa dos 1900 nm para o final do espectro analisado a curva deste material fica mais próxima da curva da variedade IACSP 96-2042.

Já as curvas médias da primeira derivada das duas coletas juntas destacam-se na região do VIS, nas feições centralizadas em 523 nm (Figuras 5 B). Destacando a região do red-edge, por volta dos 726 nm, a primeira derivada



ressaltou as diferenças das curvas médias nesta faixa, algo que não era possível perceber visualizando as curvas brutas (Figuras 5 B).

Já nas regiões correlacionadas com o conteúdo de água foliar no infravermelho de ondas curtas verifica-se pontos mínimos dos vales em 1393 nm e em 1878 nm (Figuras 5 B).

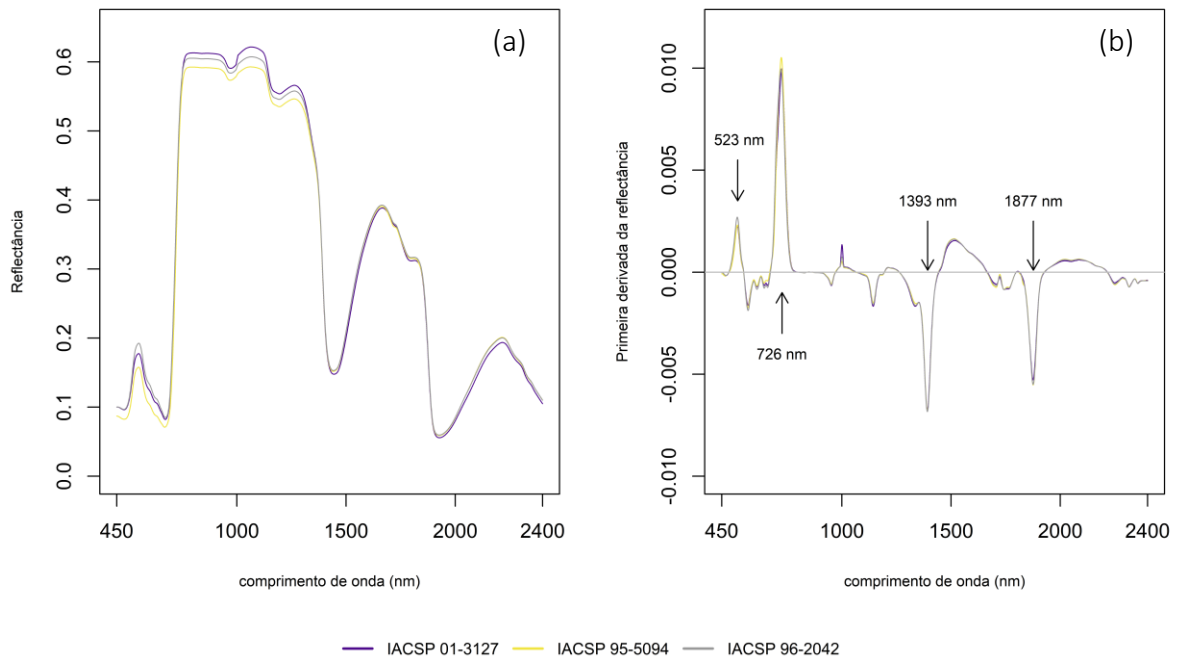


**Figura 6.** Gráfico com as curvas espectrais médias e da análise de componentes principais para as variedades IACSP 01-3127, IACSP 95-5094 e IACSP 96-2042 das duas coletas. (A) Curvas hiperespectrais brutas das variedades; (B) Curvas da 1ª derivada.

Na Coleta 1, na região do visível, a variedade IACSP 95-5094 absorveu maior quantidade de energia em relação aos demais materiais. Já a variedade IACSP 96-2042 apresentou os maiores valores de reflectância, principalmente na região do verde ao redor de 550 nm, onde há uma maior separação das curvas médias em relação ao restante do VIS (Figura 6 A). No infravermelho próximo observa-se que a variedade IACSP 01-3127 possui as maiores reflectâncias, com intensidades próximas a IACSP 96-2046 e a IACSP 95-5094, com as menores reflectâncias.

Na região do SWIR a partir 1400 nm, onde é possível observar uma divergência das curvas médias (Figura 6 A), observa-se a IACSP 01-3127 com os menores valores nesta localidade do espectro, principalmente na faixa de 1900 nm até os 2400 nm. Ressalta-se valores médios muito similares entre as três variedades até o final do espectro analisado, principalmente entre as variedades IACSP 95-5094 e IACSP 96-2042, mas nas feições centralizadas em 1400 nm e 1900 nm, há uma certa separabilidade destas curvas.

As curvas médias da primeira derivada destacam-se na região do VIS, com feição centrada em 523 nm (Figuras 6 B). No NIR na região do red-edge, observa-se pico ao redor dos 726 nm (Figuras 6 B) e no infravermelho de ondas curtas pontos mínimos em 1393 nm e em 1877 nm (Figuras 6 B), sendo evidente neste último a maior divergência das curvas entre variedades.



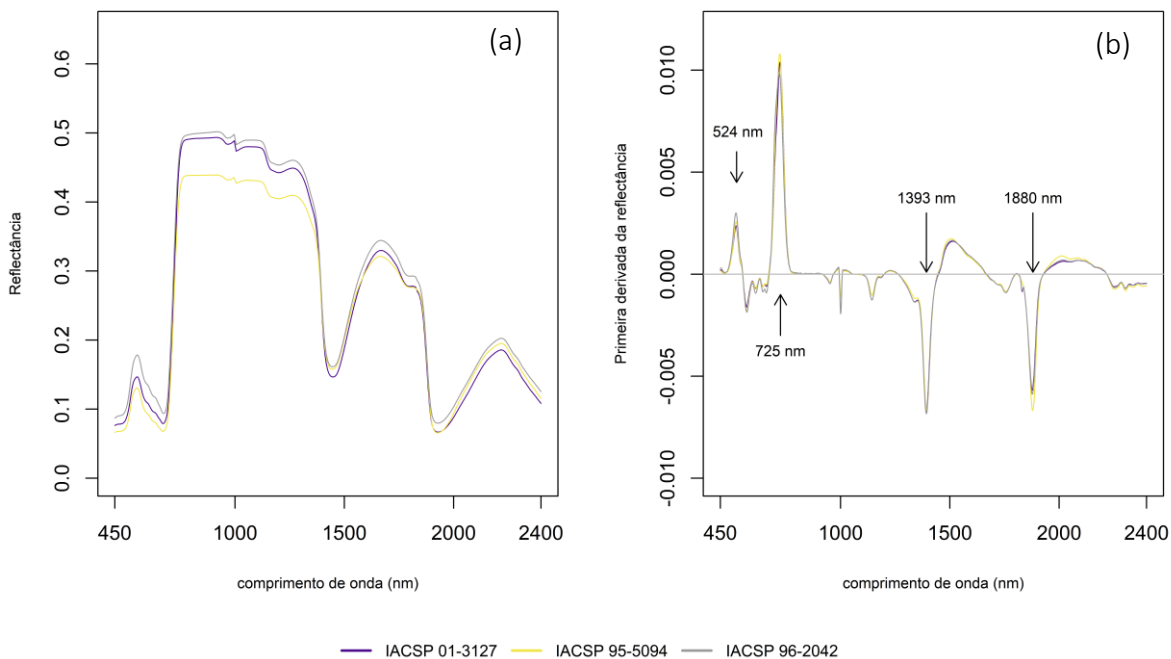
**Figura 7.** Gráfico com as curvas espectrais médias e da análise de componentes principais para as variedades IACSP 01-3127, IACSP 95-5094 e IACSP 96-2042 da Coleta 1. (A) Curvas hiperespectrais brutas das variedades; (B) Curvas da 1ª derivada.

Na coleta 2, na região do visível e do infravermelho próximo, a variedade IACSP 95-5094 absorveu maior quantidade de energia nestas faixas em relação à IACSP 96-2042. Logo, ela obteve os menores valores de reflectância, principalmente na região do verde com feição central em 550 nm (Figura 7 A). A curva média da variedade IACSP 01-3127 encontra-se mais equidistante das demais curvas, e observa-se uma melhor separação das curvas médias para esta coleta.

Para a faixa do SWIR na feição centrada em 1400 nm, a IACSP 95-5094 possui valores intermediários, quase se igualando a IACSP 96-2042 e a IACSP 01-3127 com maior absorção. Em 1900 nm, a variedade IACSP 95-5094 possui valores inferiores às demais, porém mais parecidos com a variedade IACSP 01-3127.

As curvas médias da primeira derivada destacaram à região do verde no VIS, formando feições centralizadas próximas aos 524 nm nas curvas médias (Figura 7 B), no NIR destaca-se a localização do red-edge com pico por volta dos 725 nm, destacando a variedade IACSP 95-5094 com valores superiores em relação as outras duas (Figura 7 B).

Já nas regiões correlacionadas com o conteúdo de água foliar no infravermelho de ondas curtas, verifica-se pontos médios mínimos nas feições em 1393 nm e 1880 nm (Figuras 7 B), verifica-se que nesta última feição a variedade 95-5094 se diferencia dos demais materiais, com os menores valores.



**Figura 8.** Gráfico com as curvas espectrais médias e da análise de componentes principais para as variedades IACSP 01-3127, IACSP 95-5094 e IACSP 96-2042 da Coleta 2. Curvas hiperespectrais brutas das variedades; (B) Curvas da 1ª derivada.

Comparando as duas coletas, parece haver uma maior similaridade da variedade IACSP 95-5094 com a IACSP 01-3127 na região do visível e do SWIR, enquanto infravermelho próximo a IACSP 01-3127 e IACSP 96-2042 possuem intensidades mais semelhantes (Figura 6 e 7).

A faixa do VIS está correlacionada principalmente aos pigmentos fotossintéticos, como carotenoides, xantofilas e clorofilas (445 e 645 nm), este último corresponde às faixas do azul e vermelho onde ocorre maior absorção da energia luminosa, enquanto na região do verde (550 nm) uma maior reflectância (Ponzoni, 2012).

Nas duas coletas observam-se as menores reflectâncias para a variedade IACSP 95-5094, seguida da IACSP 01-2731 (Figura 6A e 7A). Já a variedade IACSP 96-2042 se diferencia das duas anteriores com intensidades maiores nesta faixa. Segundo Clevers e Gitelson (2013) e Cui e Zhou (2017), a reflectância desta região está ligada aos teores de clorofila e correlacionam teores baixos do pigmento a reflectâncias com maiores intensidades.

Observa-se que a região do verde das curvas brutas em ambas as coletas é onde mais as curvas se divergem na região do visível, sendo que esta diferença também é observada na região do vermelho, porém com menor pujança. Segundo Peñuelas e Filella (1998), a região do verde ao redor dos 550 nm está correlacionada aos teores foliares de clorofila. Tal fato demonstra que as diferenças encontradas nas curvas se relacionam aos teores de clorofila nas folhas.

Já na 1ª derivada, para a região ao redor dos 531 nm, o local não tem muito destaque nas curvas brutas, porém ganha importância nas curvas de 1ª derivada, com as maiores feições para o VIS. Este local é associado a xantofilas, pigmentos relacionados à fotoproteção (INOUE; DARVISHZADEH; SKIDMORE, 2019; PEÑUELAS; FILELLA, 1998). Observa-se uma pequena diferença entre as coletas para esta região. Os picos deslocam-se de 523 nm na primeira coleta para 524 nm na segunda coleta, além de diferentes intensidades para as variedades na coleta 1 e mesmos valores para todas variedades na coleta 2 (Figura 6 B e 7 B).

A região do NIR está ligada à organização celular do mesófilo e seu conteúdo de água e ar (Ponzoni 2012), nesta região observou-se maior proximidade das variedades IACSP 96-2042 e IACSP 01-3127, porém observa-se na coleta 1 nesta região a variedade IACSP 01-3127 com as maiores reflectâncias e na coleta 2 a IACSP 96-2042 com as maiores reflectâncias. Tal fato indica maiores similaridades na organização celular da epiderme e mesófilo foliar entre si, enquanto a IACSP 95-5094 se afasta nitidamente delas (Figura 7A).

Fiorio et al. (2018) indica mudanças na intensidade de reflectância nas feições ao redor de 970 nm e 1200 nm e associam ao conteúdo de água das folhas. Sabe-se que a disposição das paredes celulares além quantidade de água das células e quantidade de ar no mesófilo foliar influem na resposta espectral nesta região (PONZONI, 2012). Observa-se que mesmo com diferentes acumulados de chuva observados entre as coletas, todas as variedades se mantiveram mais similares, principalmente na coleta 1 onde a restrição hídrica foi maior (Figura 6 A), e aparecendo praticamente iguais nas curvas médias das coletas juntas (Figura 5 A). Quando se trata das curvas médias da 1ª derivada para esta região, há pouca expressividade e valores bem próximos de zero, não contribuindo muito na separação das variedades. Contudo, a região do red-edge é o local de maior destaque ao longo do espectro analisado, abrangendo a faixa de 680 a 800 nm (JENSEN, 2011). O red-edge está mais ligado a questões como nitrogênio foliar e conseqüentemente a quantidade de clorofila nas plantas (BARROS et al., 2022; MARTINS et al., 2021; ZHAO et al., 2005). Martins et al. (2021) também utilizaram a primeira derivada para cana-de-açúcar para predição do teor de nitrogênio foliar. Ressalta-se que no intuito de discriminar as variedades, essa região possui grande importância.

Além do nitrogênio, Gitelson; Merzlyak; Lichtenthaler (1996) verificam alterações no posicionamento do pico do red-edge na 1ª derivada em relação a diferentes concentrações de clorofila nas folhas de *Aesculus hippocastanum* L e *Acer spp.*, deslocando dos 685 nm até os 705 nm conforme o aumento dos teores de clorofila nas folhas, em concordância com os trabalhos de Martins et al. (2021) e Filella et al. (1995). No presente trabalho, verificou-se uma pequena mudança no ponto central do pico da coleta 1 para coleta 2 para a variedade IACSP 01-3127 e IACSP 96-2042, com deslocamento de 1 nm para baixo, já a variedade IACSP 95-5094 não apresentou deslocamento algum do ponto central do red-edge.

Já o SWIR foi a região onde as variedades mais apresentavam diferenças em suas intensidades, principalmente nos vales ao redor de 1400 nm e 1900 nm, sendo que a 1ª derivada destacou bem estas regiões. Segundo Peñuelas e Filella (1998) a região o SWIR na faixa de 1300 nm a 2500 nm possui alta sensibilidade aos teores de água na folha, principalmente nas faixas ao redor de 1400 nm e 1900 nm, o que indica que neste momento as plantas encontravam-se com teores de água nas folhas semelhantes uma das outras.

A água parece estar muito ligada com mudanças observadas entre as coletas, verifica-se que as intensidades de reflectância das curvas são inferiores na coleta 2 (Figura 6 A) em relação a coleta 1 (Figura 7 A), isto está ligado à ocorrência de período mais seco no momento da coleta 1, enquanto no momento da coleta 2 o solo encontrava-se úmido devido às precipitações ocorridas, conforme fica demonstrado na Figura 2. FIORIO et al. (2018) e CHEN et al., (2012) observaram aumentos nas intensidades de reflectância na região do visível e infravermelho próximo em folhas de cana-de-açúcar após a diminuição da umidade do solo. Por sua vez, Inoue, Darvishzadeh e Skidmore (2019), estudando folhas de amendoim, correlacionaram a eficiência da fotossíntese com o conteúdo de água nas folhas desta leguminosa, além de verificarem um aumento da reflectância ao longo de todo espectro em situações de deficiência hídrica.

Diante disso, é possível afirmar que as mudanças de intensidade foram principalmente devido às diferenças do conteúdo de água nas folhas, que indiretamente afetam a região do VIS causadas pelos acumulados de

precipitações em cada coleta, o que dificulta a discriminação das variedades quando se utiliza amostras coletadas em datas distintas.

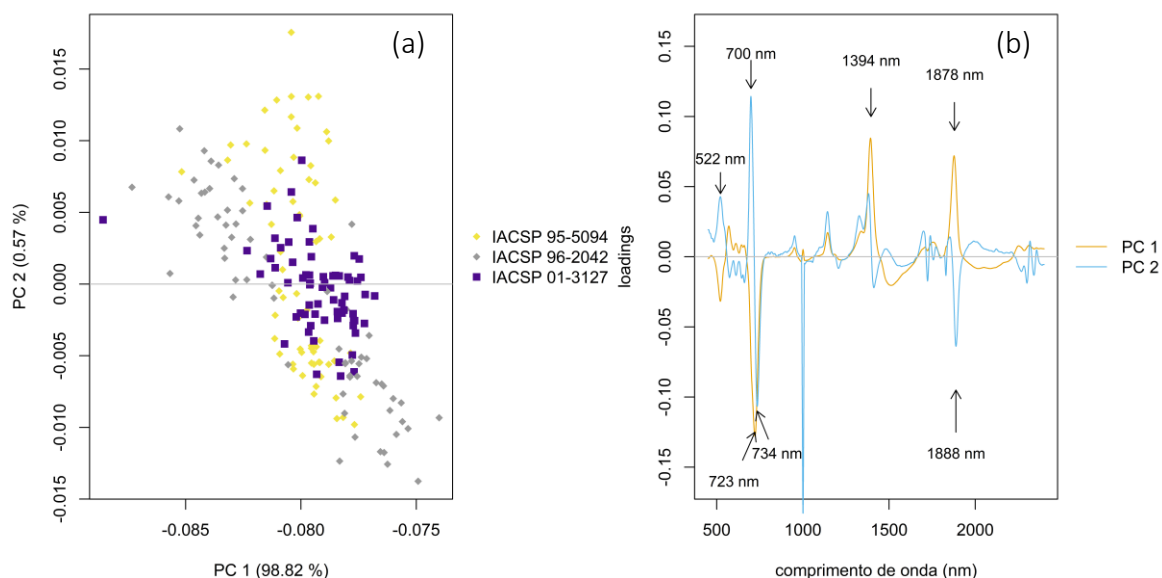
Mas as diferenças observadas no VIS e red-edge corroboram com a situação observada em campo. Verifica-se que os blocos selecionados possuem diferentes níveis de falhas e fechamentos das linhas (Figura 1), sendo a cultivar IACSP 95-5094 aquela com melhor resposta às condições ambientais locais, enquanto a IACSP 96-2042 teve pior resposta com maior número de falhas e dossel mais aberto em relação aos outros blocos. Isto corrobora com as curvas mostrando que a variedade com a maior eficiência fotossintética proporcionou um melhor fechamento do dossel e redução de falhas nas linhas de plantio.

Conforme a Tabela 1 a variedade IACSP 96-2042 é a mais exigente dos materiais selecionados, seguida das IACSP 01-3127 e IACSP 95-5094, a rusticidade da cultivar SP 95-5094 pode ter proporcionado uma maior adaptação as condições do plantio até a segunda coleta, como observado nas curvas espectrais médias (Figura 6 A e 7 A) e na imagem antes do fechamento do Dossel (Figura 1). Déficits hídricos estão associados a mudanças morfofisiológicas e interferências no desenvolvimento dos vegetais (CHAVES; FLEXAS; PINHEIROS, 2009), podendo ser o fator que levou a diferenciação encontradas nesta faixa do espectro quando às condições hídricas se normalizaram, ou seja, corroborando com resultados obtidos por este trabalho.

### **2.3.3. Análise das Componentes Principais (PCA)**

Utilizou-se a análise de componentes principais para observar o comportamento das variedades de cana-de-açúcar selecionadas. O gráfico de dispersão dos scores (Figura 8 A) utilizando as componentes 1 e 2 (98,57% e 0,72% da variância observada, respectivamente), conseguem separar as variedades, porém os materiais IACSP 01-3127 e a IACSP 95-5094 se sobrepõem entre si. Observa-se que a componente principal 1 é a responsável pela diferença entre estas variedades, enquanto a componente principal 2 é responsável por um espalhamento dos pontos de cada material ao longo do eixo y (Figura 5 A e B) indicando que a variabilidade entre as coletas, como é observado com a variedade IACSP 96-2042 que parece formar 2 subgrupos.

Os loadings para as coletas juntas (Figura 8 B) mostram para a componente principal 1 um maior peso para o red-edge com um vale centrado em 723 nm, dois picos centrados nas regiões de influência do conteúdo de água nas folhas (1394 nm e 1878 nm) e com menor influência que os demais um vale centrado em 522 nm na região do verde. Para a componente principal 2 há um pico (700 nm) e um vale (734 nm) no red-edge, e um vale em 1888 nm, e com menor importância um pico na região de 522 nm e 1394 nm.

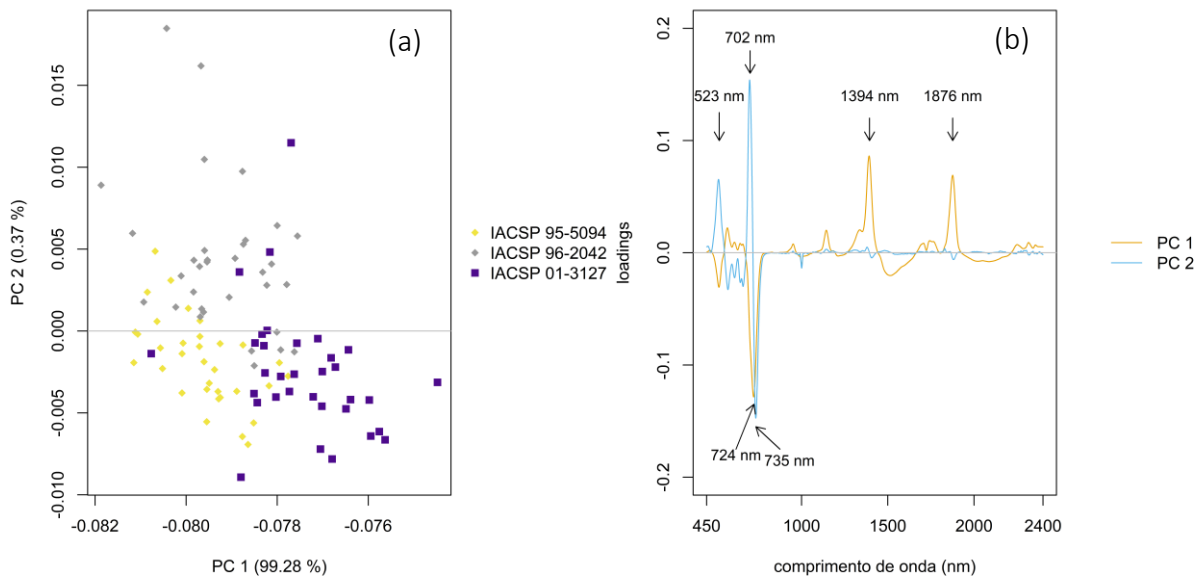


**Figura 9.** Gráfico da análise de componentes principais para as duas coletas; (A) gráfico de dispersão dos scores das componentes principais 1 e 2; (B) loadings das componentes principais 1 e 2.

Já o gráfico dos scores para a coleta 1 (Figura 9 A) com as componentes principais 1 e 2, que representam 99,14% e 0,38% da variância observada, demonstra uma excelente separabilidade das cultivares, porém ainda são verificadas sobreposições entre todas as variedades, principalmente entre os materiais IACSP 95-5094 e IACSP 96-2042. Nesta coleta observa-se a que na região do infravermelho de ondas curtas, estas duas cultivares possuem curvas médias com valores mais próximos (Figura 6 A).

Os loadings da primeira coleta (Figura 9 B) mostram para a componente principal 1 um maior peso para o red-edge com um vale centrado em 724 nm, dois picos centrados nas regiões de influência do conteúdo de água nas folhas (1394 nm e 1876 nm) e um vale com menor influência que os demais ao redor de 523 nm na região do verde. Para a componente principal 2 há um pico (702 nm) e um vale (735 nm) no red-edge, e com menor importância um pico na região de 523 nm.

A componente principal 1 é responsável pela separação dos materiais entre si, porém não consegue separar muito bem as variedades IACSP 96-2042 da IACSP 01-3127, sendo a componente principal 2 a responsável por isto. Enquanto a componente 1 possui uma maior influência das bandas na faixa do 1400 nm e 1900 nm, a componente 2 está ligada ao visível, região associada a mecanismos de fotoproteção e o red-edge ligado aos teores de nitrogênio foliar, e nenhuma influência das bandas ligadas ao teor de água das folhas, corroborando com as restrições hídricas que afetavam cultura naquele momento.

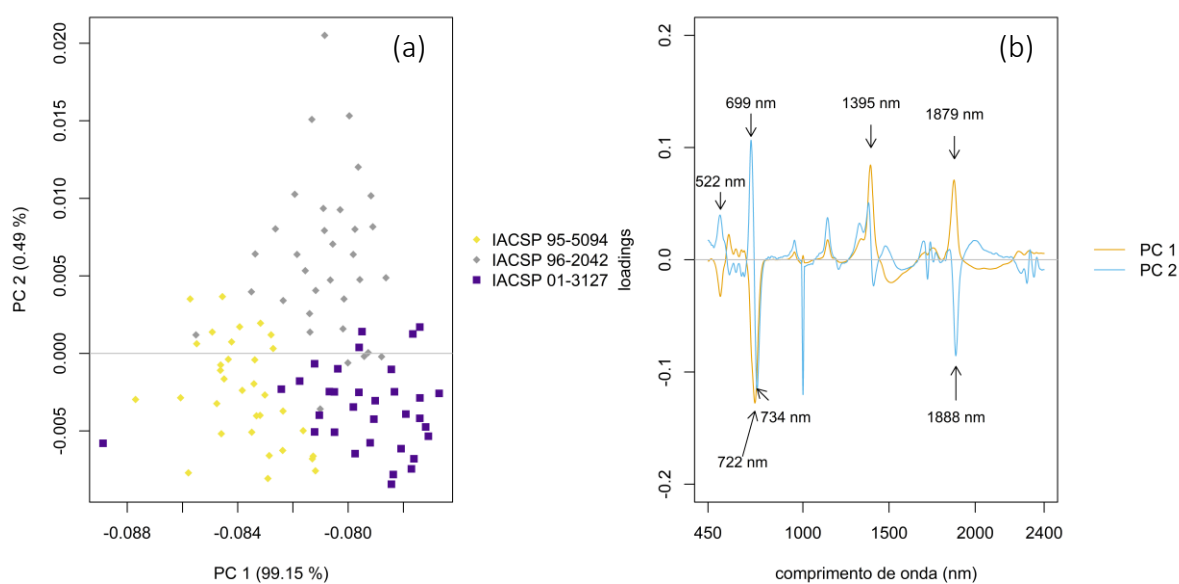


**Figura 10.** Gráfico da análise de componentes principais para primeira coleta; (A) gráfico de dispersão dos scores das componentes principais 1 e 2; (B) loadings das componentes principais 1 e 2.

Já a coleta 2 o gráfico dos scores (Figura 10 A) para as componentes principais 1 e 2 (variâncias de 99,06% e 0,51% respectivamente), observa-se um melhor agrupamento com menores sobreposições destes grupos. Nesta coleta nota-se uma maior sobreposição entre a variedade 96-2042 com a 01-3127, mas de maneira tênue.

Os loadings para a coleta 2 (Figura 10 B) mostram para a componente principal 1 um maior peso para o red-edge com um vale centrado em 722 nm, dois picos centrados nas regiões de influência do conteúdo de água nas folhas (1394 nm e 1879 nm) e influência muito reduzida da faixa ao redor dos 522 nm na região do verde. Já a componente principal 2 possui um pico em 699 nm e um vale 734 nm no red-edge, e um vale em 1888 nm, e com menor importância um pico na região de 522 nm e 1394 nm.

A coleta 2 apresentou os melhores agrupamentos das variedades, com pouca sobreposição entre os materiais em relação a coleta 1 (Figura 9 A), sendo a coleta que apresentou a maior influência do conteúdo de água foliar nas componentes 1 e 2. Observa-se que a componente principal 1 consegue separar apenas as variedades IACSP 01-3127 e IACSP 95-5094, enquanto a componente principal 2 separa a IACSP 96-2042 das demais coletas. Ressalta-se que nesta coleta foi inexpressiva a região do visível em ambas as componentes (522 nm), este local teve redução de seu peso devido as condições hídricas terem voltado a normalidade.



**Figura 11.** Gráfico da análise de componentes principais para segunda coleta; (A) gráfico de dispersão dos scores das componentes principais 1 e 2; (B) loadings das componentes principais 1 e 2.

É observado que apesar da pequena influência, a região do VIS possui uma estabilidade maior nas regiões de seus picos e vales em ambas as coletas, o que não ocorre com as bandas de água. Verifica-se também que na coleta 2, a componente principal 2 possui um vale na região de 1888 nm não observado na coleta 1, vindo a influenciar o gráfico de scores, diminuindo a sobreposição das variedades de cana em relação a coleta 1 que este vale é ausente na componente principal 2.

Os *loadings* das componentes principais 1 e 2 (Figura 8 B,9 B e 10 B) possuem grande influência do conteúdo de água foliar, com picos localizados em 1394 nm e 1878 nm na componente principal 1 e um vale com valor mínimo em 1888 nm na componente 2. Já na região red-edge com vales por volta dos 700 nm e 723 nm e picos por volta dos 734 nm, quando tratamos as coletas de forma separada a região do verde ganha um maior peso nas componentes principais principalmente na componente 2 ao redor da banda dos 523 nm. A maior diferença é observada na coleta 1 (Figura 9 B) com os loadings da componente 2, onde verifica-se um vale na região de influência da água ao redor dos 1888 nm, o que não ocorre na coleta 2 (Figura 10 B). O que também ocorreu com a região do red-edge. Oliveira et al. (2020) também observou em seu estudo com algodão, grande influência dos *loadings* na região do red-edge e as bandas da água na faixa dos 1400 nm e 1900 nm.

As demais componentes não possibilitaram bons agrupamentos, e como consequência houve maiores sobreposições das variedades, motivo por não terem sido plotados gráfico dos scores.

Ressalta-se aqui que o aparecimento de um ruído próximo dos 1100 nm visualizado principalmente na coleta 2 (Figura 8 B e 10 B) foi devido a mudanças dos sensores do NIR para o sensor na faixa do SWIR. Apesar de aparecerem valores elevados nos *loadings* da segunda componente e também os da primeira derivada, eles estão concentrados em uma pequena faixa, e, portanto, apresentam uma pequena contribuição na formação dos valores da segunda componente principal da coleta 2, além dos baixos valores da 1ª derivada (Figura 10 B e 8 B).

Martins et al. (2021) utilizando a técnica estatística sparse partial least square (SPLS) para determinar as bandas com maior relevância para a predição de teores de nitrogênio foliar também verificaram importância na região do verde, com feição centralizada em 525 nm e red-edge, entre os 680 e 715 nm. Já Stedile Neto et al. (2018) em seu estudo utilizando colmos para avaliar o potencial deste tecido na classificação de variedades de cana



verificaram que a região de 600 nm a 750 nm tiveram os maiores pesos utilizando 3 tipos de análise multivariadas, corroborando os resultados deste trabalho.

Fica evidente que o fator água é o de maior peso no agrupamento das variedades e as discriminação entre elas, junto com as regiões do red-edge e região do verde na área de influência das Xantofilas, que são indiretamente relacionadas a questões hídricas das plantas (FIORIO et al., 2018; CHEN et al., 2012). Os fatores que mais afetam a discriminação de plantas são biofísicos, estágios de desenvolvimento, manejo e aspectos locais (GALVÃO et al, 2011), desta forma o experimento reduziu a influência de alguns destes fatores com plantio na mesma época, mesmo manejo de adubação e controle de plantas daninhas, pragas e doenças, mesmo solo e plantio próximo. Assim, as diferenças na resposta espectral decorrem de fatores fisiológicos e morfológicos de cada variedade e sua resposta diferenciada as restrições e acesso nutrientes, água e energia solar, ajudando na separação de cada material nos gráficos dos scores, porém para as componentes principais 1 e 2 está muito ligado na eficiência na utilização da água disponível no ambiente.

### 2.3.4. Classificação das variedades de Cana-de-Açúcar

No intuito de verificar a possibilidade de classificação das variedades selecionadas, pelo fato da maioria dos espectrômetros de campo utilizarem o VIS-NIR de 350 nm até os 1100 nm na coleta de dados, além da grande influência das bandas relacionadas ao conteúdo de água nas folhas, utilizou-se a faixa visível e região do red-edge segundo definição de Jensen (2011), abrangendo os 450 a 800 nm. A utilização desta faixa buscou atenuar a influência d'água, logo que os pesos para as bandas ao redor de 1400 nm e 1900 nm possuem valores altos e influenciam os valores das componentes, mascarando outras feições das curvas que possam estar ligadas a características intrínsecas a cada variedade.

**Tabela 2.** Matriz de confusão da classificação das variedades de Cana-de-Açúcar com a utilização da aprendizagem de máquinas Random Forest utilizando a coleta 1 como amostra de treinamento do modelo.

Classificados\Referências	IACSP 01-3127	IACSP 95-5094	IACSP 96-2042
IACSP 01-3127	19	1	4
IACSP 95-5094	7	28	5
IACSP 96-2042	7	4	24

Para manter a independência das amostras para os grupos de treinamento do modelo e o grupo de predição utilizou-se as coletas. Na tabela 1 estão os resultados com a utilização da coleta 1 como base do modelo de predição e na tabela 2 os resultados utilizando a coleta 2 como base do modelo.

A classificação utilizando como base para o modelo os dados da coleta 1 obteve Acurácia Global de 71,72% e índice Kappa de 0,57 enquanto a utilização da coleta 2 como base para o treinamento do modelo obteve acurácia global de 67,68% e índice Kappa de 0,5.

**Tabela 3.** Matriz de confusão da classificação das variedades de Cana-de-Açúcar com a utilização da aprendizagem de máquinas Random Forest utilizando a coleta 2 como amostra de treinamento do modelo.

Classificados\Referências	IACSP 01-3127	IACSP 95-5094	IACSP 96-2042
IACSP 01-3127	23	1	3
IACSP 95-5094	4	32	19
IACSP 96-2042	6	0	11

Pode se considerar que os resultados da classificação foram moderados. A variedade IACSP 95-5094 obteve os melhores resultados com acurácia do produtor de 84,85% e 96,97%, e acurácia do usuário de 70% e 58,18%, seguido da IACSP 01-3127 com acurácia do produtor de 57,58 % e 69,70%, e acurácia do usuário de 85,18% e 79%, e por último a IACSP 96-2042 com acurácia do produtor de 72,73% e 33,33% e acurácia do usuário de 68,57% e 64,7%, tendo maior número de amostras classificadas como a variedade 95-5094.

Utilizando-se a coleta 2 como treinamento do modelo, observou-se péssimo resultado para a variedade 96-2042, com apenas 33,33% de acurácia do produtor.

Estudo realizado em casa de vegetação com colmos de 4 variedades diferentes de cana-de-açúcar tiveram bons resultados na classificação, obtendo 82% de classificações corretas utilizando a técnica PLS-DA. Porém, quando se verifica a classificação das variedades mais semelhantes entre si, ocorre grande confusão na classificação, com valores inferiores a 50% entre 2 cultivares (STEDILE NETO et al., 2018).

Galvão, Formaggio e Tisot (2005) tiveram bons resultado em discriminar 5 variedades de cana-de-açúcar utilizando o satélite Hyperion com uma acurácia global de 87,5%, porém características do dossel das plantas vem a contribuir com a discriminação, como o fato de algumas variedades possuírem dossel planófilo e outras terem erectófilo, características que interferem na resposta espectral, além de diferenças entre os solos de cada pixel utilizado no modelo de discriminação canônica.

## 2.4. Conclusões

É possível separar variedades de cana-de-açúcar utilizando dados hiperespectrais de folhas.

As regiões do espectro ligadas ao red-edge e as bandas correlacionadas ao conteúdo de água nas folhas são as regiões que vem a contribuir com a separação das variedades, influenciando as componentes com as maiores variâncias.

O uso do algoritmo Random Forest utilizando-se as 5 componentes principais teve bons resultados, empregando dados de coletas de épocas diferentes.

Os dados hiperespectrais corroboraram a adaptabilidade das variedades às condições ambientais enfrentadas por estas na safra em que as coletas aconteceram.

## Agradecimentos

O presente trabalho foi realizado em parte com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Contando também com apoio da FAPESP pela aquisição do equipamento FieldSpec® 3 Jr através do projeto de pesquisa no 2013/22435-9.

## Referências

- ACEVEDO, F.J.; JIMÉNEZ, J.; MALDONADO, S.; DOMÍNGUEZ, E.; NARVÁEZ, A. Classification of wines produced in specific regions by UV-visible spectroscopy combined with support vector machines. **Journal of Agricultural and Food Chemistry**, v.55, p.6842–6849, 2007. Disponível em: <<https://doi.org/10.1021/jf070634q>>.
- ADAMS, M.J. **Chemometrics in analytical spectroscopy**. Wolverhampton: Bookcraft, 1995.
- ALMEIDA, M. de.; CROCOMO, O. J. Organografia de Dez Cultivares de Cana-de-açúcar (*Saccharum spp.*), I. **Revista de Agricultura**. Piracicaba, v. 69, n.1, p.41-65, 1994.
- ALMEIDA, M. de; ROCHELLE, L.A.; CROCOMO, O. J. Chave analítica para determinação de dez variedades de cana-de-açúcar (*Saccharum spp.*). **Scientia Agricola**. Piracicaba, v.52, n. 1, p. 16-19, jan./abr., 1995.
- ASD – Analytical Spectral Devices. FieldSpec® FieldSpec 3 User Manual. Boulder CO, 2010. 110 p.
- BAO, D., ZHOU, J., BHUIVAN, S. A., ZIA, A., FORD, R., & GAO, Y. Early Detection of Sugarcane Smut Disease in Hyperspectral Images. International Conference on Image and Vision Computing, 36, 2021, Tauranga, IEEE Xplore. Tauranga: IEEE, 2021.p. 1–6. Disponível em: <<https://doi.org/10.1109/IVCNZ54163.2021.9653386>>.
- BARROS, P. P. da S., FIORIO, P. R., DEMATTÊ, J. A. de M., MARTINS, J. A., MONTEZANO, Z. F., & DIAS, F. L. F. (2022). Estimation of leaf nitrogen levels in sugarcane using hyperspectral models. **Ciência Rural**, n.52, v.7, 2022. Disponível em: <<https://doi.org/10.1590/0103-8478cr20200630>>.
- CHAVES, M.M.; FLEXAS, J.; PINHEIRO, C. Photosynthesis under drought and salt stress : regulation mechanisms from whole plant to cell. **Annals of Botany**, v. 4, n. 104, p.551–560, 2009. Disponível em: <<https://doi.org/10.1093/aob/mcn125>>.
- CHEN, H.; WANG, P.; LI, J.; ZHANG, J.; ZHONG, L. Canopy Spectral Reflectance Feature and Leaf Water Potential of Sugarcane Inversion. **Physics Procedia**, v.25, p.595–600, 2012. Disponível em: <<https://dx.doi.org/10.1016/j.phpro.2012.03.131>>
- CHEAVEGATTI-GIANOTTO, A.; de ABREU A. P.; BESPALHOK FILHO, J. C.; BURNQUIST, W. L.; CRESTE, S; di CIERO; L.; FERRO, J. A., de Oliveira Figueira A. V.; de SOUZA FILGUEIRAS T.; GROSSI-DE-SÁ M. D.; GUZZO E. C., HOFFMANN H. P., de ANDRADE LANDELL M. G.; MACEDO N.; MATSUOKA S.; de CASTRO REINACH F.; ROMANO E.; da SILVA W. J., de CASTRO SILVA FILHO M.; CESAR ULIAN E. Sugarcane (*Saccharum X officinarum*): A Reference Study for the Regulation of Genetically Modified Cultivars in Brazil. **Trop Plant Biol. Tropical Plant Biology**, v. 4, n. 1, p. 62 - 89, 2011. Disponível em:< <https://dx.doi.org/10.1007%2Fs12042-011-9068-3>>
- CORTES, C; VAPNIK, V. Support-Vector Networks. **Machine Learning**, v. 20, mai., p. 273-297, 1995.
- CLEVERS, J.G.P.W.; GITELSON, A.A. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 23, n. 1, p. 334–343, 2013.
- CUI, S.; ZHOU, K. A comparison of the predictive potential of various vegetation indices for leaf chlorophyll content. **Earth Science Informatics**, Heidelberg, v. 10, n. 2, p. 169–181, 2017. Disponível em: <<https://doi.org/10.1007/s12145-016-0281-3>>
- CUTLER, D. R.; EDWARDS Jr., T. C.; BEARD, K. H.; CUTLER, A.; HESS, K. T.; GIBSON, J.; LAWLER, J. J. Random Forests for classification in ecology. **Ecology**. n.88, p. 2783-2792, 2007.

- DAO, P. D.; HE, Y.; PROCTOR, C. Plant drought impact detection using ultra-high spatial resolution hyperspectral images and machine learning. **International Journal of Applied Earth Observation and Geoinformation**, n.1, v.102, p.1-18, 2021. Disponível em: <<https://doi.org/10.1016/j.jag.2021.102364>>.
- DOE. **Alternative Fuels Data Center**. Global Ethanol Production. Disponível em: <<https://afdc.energy.gov/data/?q=ethanol>>. Acesso em: 15 janeiro 2021.
- ELAVARASAN, D.; VINCENT, D.R.; SHARMA, V.; ZOMAYA, A.Y.; SRINIVASAN, K. Forecasting yield by integrating agrarian factors and machine learning models: A survey. **Computers and Electronics in Agriculture**, v.155, p.257–282, 2018. Disponível em: <<https://doi.org/10.1016/j.compag.2018.10.024>>.
- ESPINOSA, C. E.; VELASQUEZ, S.; HERNANDEZ, F. L. Sugarcane Productivity Estimation through Processing Hyperspectral Signatures Using Artificial Neural Networks. In: IEEE Latin American GRSS and ISPRS Remote Sensing Conference, LAGIRS, 2020, Santiago. 2020 IEEE Latin American GRSS ISPRS Remote Sensing Conference (LAGIRS), IEEE Xplore: Santiago, 2020, p.290–295. Disponível em: <<https://doi.org/10.1109/LAGIRS48042.2020.9165683>>. FAO - Food and Agriculture Organization. Statistics Division. **Area harvested, yield and sugarcane production in the world**. Disponível em: <<http://www.fao.org/faostat/en/#data/QC>>. Acesso em: 15 jan. 2021.
- FABIYI, S.D.; VU, H.; TACHTATZIS, C.; MURRAY, P.; HARLE, D.; DAO, T.; ANDONOVIC, I.; REN, J.; MARSHALL, S. Comparative Study of PCA and LDA for Rice Seeds Quality Inspection. In: 2019 IEEE AFRICON, 2019. ACCRA: IEEE, 2019, p. 1-4. Disponível em: <<https://doi.org/10.1109/AFRICON46755.2019.9134059>>.
- FILELLA, I; SERRANO, L.; SERRA, J.; PEÑUELAS. Evaluating Wheat Nitrogen Status with Canopy Reflectance Indices and Discriminant Analysis. *Crop Science*, v. 35, n. 5, p. 1400-1405, 1995. Disponível em: <<https://doi.org/10.2135/cropsci1995.0011183X003500050023x>>.
- FIORIO, P.R.; COELHO, R.D.; BARROS, P.P.S.; BONILLA, M.M.Z.; GADY, A.P.B. Comportamento Espectral De Folhas Da Cana-De-Açúcar Na Presença De Déficit Hídrico. **Irriga**, v.23, p.609–621, 2018. Disponível em: <<http://dx.doi.org/10.15809/irriga.2018v23n3p609-621>>.
- FRANCISCO, A. C. **O regime da propriedade intelectual para a Cana-de-Açúcar**. Tese. Faculdade de Direito, Universidade de São Paulo, São Paulo. 2014.
- FRASSON, F.R.; MOLIN, J.P.; POVH, F.P.; SALVI, J.V. Temporal Behavior of Ndvi Measured With an Active Optical Sensor for Different Varieties of Sugarcane / Comportamento Temporal Do Ndvi Mensurado Com Sensor Óptico Ativo Para Diferentes Variedades De Cana-De-Açúcar. **Revista Brasileira de Engenharia de Biosistemas**, v.1, p.237, 2007. Disponível em: <<https://doi.org/10.18011/bioeng2007v1n3p237-244>>.
- GALVÃO, L.S.; FORMAGGIO, A.R.; TISOT, D.A. Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data. **Remote Sensing of Environment**, v.94, p.523–534, 2005. Disponível em: <<https://doi.org/10.1016/j.rse.2004.11.012>>.
- GALVÃO, L. S.; EPIPHANIO, J. C. N.; BREUNING, F. M.; FORMAGGIO, A. R. **Crop Type Discrimination Using Hyperspectral Data**. Hyperspectral remote sensing of vegetation. CRC Press, p. 329–358, New York, 2011.
- GAO, D.; LI, M.; ZHANG, J.; SONG, D.; SUN, H.; QIAO, L.; ZHAO, R. Improvement of chlorophyll content estimation on maize leaf by vein removal in hyperspectral image. **Computers and Electronics in Agriculture**, v.184, p.106077, 2021. Disponível em: <<https://doi.org/10.1016/j.compag.2021.106077>>.

- GITELSON, A.A.; MERZLYAK, M.N.; LICHTENTHALER, H.K. Detection of Red Edge Position and Chlorophyll Content by Reflectance Measurements Near 700 nm. **Journal of Plant Physiology**, v.148, p.501–508, 1996. Disponível em: <[https://doi.org/10.1016/S0176-1617\(96\)80285-9](https://doi.org/10.1016/S0176-1617(96)80285-9)>
- GONÇALVES, M.T.V.; MOROTA, G.; ALMEIDA COSTA, P.M. DE; VIDIGAL, P.M.P.; BARBOSA, M.H.P.; PETERNELLI, L.A. Near-infrared spectroscopy outperforms genomics for predicting sugarcane feedstock quality traits. **PLOS ONE**, p.1–20, 2021. Disponível em: <<https://doi.org/10.1101/2020.07.16.206110>>.
- GRISHAM, M.P.; JOHNSON, R.M.; ZIMBA, P. V. Detecting Sugarcane yellow leaf virus infection in asymptomatic leaves with hyperspectral remote sensing and associated leaf pigment changes. **Journal of Virological Methods**, v.167, p.140–145, 2010. Disponível em: <<https://doi.org/10.1016/j.jviromet.2010.03.024>>.
- HUA, Q.; YU, Y.; DONG, S.; LI, S.; SHEN, H.; HAN, Y.; ZHANG, J. Leaf spectral responses of *Poa crymophila* to nitrogen deposition and climate change on Qinghai-Tibetan Plateau. **Agriculture, Ecosystems and Environment**, v.284, 2019. Disponível em: <<http://dx.doi.org/10.1016/j.agee.2019.106598>>.
- HUANG, Y.; LI, J.; YANG, R.; WANG, F.; LI, Y.; ZHANG, S. Hyperspectral Imaging for Identification of an Invasive Plant *Mikania micrantha* Kunth. **Frontiers of Plant Science**, v.12, p.1–12, 2021. Disponível em: <<https://doi.org/10.3389/fpls.2021.626516>>.
- Programa Cana. Campinas: IAC, [1-2], 1 folder.2012.
- Programa Cana. Campinas: IAC, [1-2], 1 folder. 2016.
- IEA-Instituto de economia agrícola. **Alta na Produção e nas Exportações de Açúcar Marca a Safra 2020/21 de Cana**. Disponível em: <<http://www.ica.sp.gov.br/out/TerTexto.php?codTexto=15925>>. Acesso em: 30 jan. 2022.
- INOUE, Y.; DARVISHZADEH, R.; SKIDMORE. **Hyperspectral Assessment of Ecophysiological Functioning for Diagnostics of Crops and Vegetation**. Biophysical and Biochemical Characterization and Plant Species Studies. CRC Press, p. 25–71, New York, 2019.
- JENSEN, J.R. **Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestre**. 3. ed. São José dos Campos: Parêntese, 2011. 598 p.
- JIE, D.; ZHOU, W.; WEI, X. Nondestructive detection of maturity of watermelon by spectral characteristic using NIR diffuse transmittance technique. **Scientia Horticulturae**, v.257, p.1-7, 2019. Disponível em: <[10.1016/j.scienta.2019.108718](https://doi.org/10.1016/j.scienta.2019.108718)>.
- LIGGES, U.; SHORT, T.; KIENZLE, P.; SCHNACKENBERG, S.; BILLINGHURST, D.; BORCHERS, H.-W.; CAREZIA, A.; DUPUIS, P.; EATON, J.W.; FARHI, E.; HABEL, K.; HORNIK, K.; KREY, S.; LASH, B.; LEISCH, F.; MERSMANN, O.; NEIS, P.; RUOHIO, J.; SMITH III, J.O.; STEWART, D.; WEINGESSEL, A. Package ‘signal’. p.1–68, 2015.
- LIU, B.; GAO, L.; LI, B.; MARCOS-MARTINEZ, R.; BRYAN, B.A. Nonparametric machine learning for mapping forest cover and exploring influential factors. **Landscape Ecology**, v.35, p.1683–1699, 2020. Disponível em: <<https://doi.org/10.1007/s10980-020-01046-0>>.
- LIU, H.; YU, T.; HU, B.; HOU, X.; ZHANG, Z.; LIU, X.; LIU, J.; WANG, X.; ZHONG, J.; TAN, Z.; XIA, S.; QIAN, B. Uav-borne hyperspectral imaging remote sensing system based on acousto-optic tunable filter for water quality monitoring. **Remote Sensing**, n.13, v.20, 2021.
- MANLY, B. J. F. **Métodos Estatísticos Multivariados: Uma Introdução**. 3º ed. Porto Alegre, Bookman, 2008, 229 p.

- MARAFON, A.C. Análise quantitativa de crescimento em cana-de-açúcar: uma introdução ao procedimento prático. Aracaju: **Embrapa Tabuleiros Costeiros**, 2012. 29p.
- MARAPHUM, K.; SAENGPRACHATANARUG, K.; APARATANA, K.; IZUMIKAWA, Y.; TAIRA, E. Spatial mapping of Brix and moisture content in sugarcane stalk using hyperspectral imaging. **Journal of Near Infrared Spectroscopy**, n. 28, v.4, p. 167–174, 2020.
- MARTINS, J.A.; FIORIO, P.R.; SILVA BARROS, P.P. DA; MELO DEMATTÊ, J.A.; MOLIN, J.P.; CANTARELLA, H.; USHER NEALE, C.M. Potential use of hyperspectral data to monitor sugarcane nitrogen status. **Acta Scientiarum - Agronomy**, v.43, p.1–13, 2021. Disponível em: <<https://doi.org/10.4025/actasciagron.v43i1.47632>>.
- MAX, A.; WING, J.; WESTON, S.; WILLIAMS, A.; KEEFER, C.; ENGELHARDT, A.; COOPER, T.; MAYER, Z.; ZIEM, A.; SCRUCICA, L.; HUNT, T.; KUHN, M.M. Package ‘caret’ R topics documented : 2020.
- MEVIK, B.-H.; WEHRENS, R. Introduction to the pls Package. **Help section of the “pls” package of RStudio software**, p.1–23, 2015.
- OLIVEIRA, M. R.; QUEIROZ, T. R. G.; TEIXEIRA, A. S.; MOREIRA, L. J.; LEÃO, R. A. O. Reflectance spectrometry applied to the analysis of nitrogen and potassium deficiency in cotton. **Revista Ciência Agrônômica**, v. 51, n. 4, p.1–10, 2020. Disponível em:< <https://doi.org/10.5935/1806-6690.20200074>>.
- PARK, H.; KIM, N.; LEE, J. Parametric models and non-parametric machine learning models for predicting option prices: Empirical comparison study over KOSPI 200 Index options. **Expert Systems with Applications**, v.41, p.5227–5237, 2014. Disponível em:< <http://dx.doi.org/10.1016/j.eswa.2014.01.032>>.
- PEÑUELAS J.; FILELLA, I. Visible and near-infrared reflectance techniques for diagnosing plant physiological status. **Trends in Plant Science**, n. 3, p.151–156, 1998.
- ROBNIK-SIKONJA, M. **Improving Random Forests**, p. 359-370, 2004. In BOULICAUT, J; ESPOSITO, F.; GIANNOTTI,F.; PEDRESCHI, D. **Machine Learning ECML 2004**, SPRINGER, Verlag Berlin Heidelberg, 2004, 580p.
- RODRIGUES, G. S. de C.; ROSS, J. L. S. **A trajetória da Cana-de-açúcar no Brasil: perspectivas geográficas histórica e ambiental**. 1 ed. Edufu, 2020. 272 p. Disponível em: <<http://doi.org/10.14393/EDUFU/978-65-86084-00-9>>
- SILALAH, D.D.; MIDI, H.; ARASAN, J.; MUSTAFA, M.S.; CALIMAN, J.P. Robust generalized multiplicative scatter correction algorithm on pretreatment of near infrared spectral data. **Vibrational Spectroscopy**, v.97, p.55–65, 2018. Disponível em:< <http://dx.doi.org/10.1016/j.vibspec.2018.05.002>>.
- SHAO, Z.; FENG, X.; BAI, L.; JIAO, H.; ZHANG, Y.; LI, D.; FAN, H.; HUANG, X.; DING, Y.; ALTAN, O.; SALEEN, N. Monitoring and Predicting Desert Locust Plague Severity in Asia-Africa Using Multisource Remote Sensing Time-Series Data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v.14, p.8638–8652, 2021. Disponível em:< <https://doi.org/10.1109/JSTARS.2021.3104936>>.
- SOUZA, M. F.; AMARAL, L. R. do; OLIVEIRA, S. R. de M.; COUTINHO, M. A. N.; FERREIRA NETTO, C. Spectral differentiation of sugarcane from weeds. **Biosystems Engineering**, n.190, p.41–46, 2020.
- STEIDLE NETO, A.J.; LOPES, D.C.; TOLEDO, J. V.; ZOLNIER, S.; SILVA, T.G.F. Classification of sugarcane varieties using visible/near infrared spectral reflectance of stalks and multivariate methods. **Journal of Agricultural Science**, v.156, p.537–546, 2018. Disponível em:< <https://doi.org/10.1017/S0021859618000539>>.

- TAVARES, T. R.; FIORIO, P. R.; SEIXAS, H. T.; GARCIA, A. C.; BARROS, P. P. S. Effects of storage on vis-NIR-SWIR reflectance spectra of Mombasa grass leaf samples. **Ciência Rural**, v. 50, n.3, 2020. Disponível em: <<https://doi.org/10.1590/0103-8478cr20190587>>.
- USDA. Foreign Agricultural Service. Sugar: World Markets and Trade, November, 2020. Disponível em <<https://apps.fas.usda.gov/psdonline/circulars/Sugar.pdf>>. Acesso em: 15 jan. 2021.
- WEHRENS, R. **Chemometrics with R. Multivariate Data Analysis in the Natural Sciences and Life Sciences**. SPRINGER, Verlag Berlin Heidelberg, 2011, 285p.
- ZHANG, Y.; WU, L.; DENG, L.; OUYANG, B. Retrieval of water quality parameters from hyperspectral images using a hybrid feedback deep factorization machine model. **Water Research**, v.204, 2021. Disponível em:<<https://doi.org/10.1016/j.watres.2021.117618>>.
- ZHAO, J.; FANG, Y.; CHU, G.; YAN, H.; HU, L.; HUANG, L. Identification of Leaf-Scale Wheat Powdery Mildew (*Blumeria graminis* f. sp. *Tritici*) Combining Hyperspectral Imaging and an SVM Classifier. **Plants**, v.9, p.1-13, 2020. Disponível em:<<https://doi.org/10.3390/plants9080936>>.
- ZHAO, D.; REDDY, K.R.; GOPAL, V.; REDDY, V.R. Nitrogen deficiency effects on plant growth, leaf photosynthesis, and hyperspectral reflectance properties of sorghum. **European Journal of Agronomy**, v.22, p.391–403, 2005. Disponível em: <<https://doi.org/10.1016/j.eja.2004.06.005>>.