

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
GOIANO - CAMPUS RIO VERDE - PROGRAMA DE PÓS-
GRADUAÇÃO EM CIÊNCIAS AGRÁRIAS - AGRONOMIA

MACHINE LEARNING PARA DETECÇÃO DE *Pratylenchus*
brachyurus NA CULTURA DA SOJA UTILIZANDO
IMAGENS MSI/SENTINEL-2

Autor: Lorena Cristina Alves Nogueira
Orientador: Prof. Dr. Alaerson Maia Geraldine
Coorientador: Prof. Dr. Tavvs Micael Alves

RIO VERDE - GO
Novembro – 2020

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
GOIANO - CAMPUS RIO VERDE - PROGRAMA DE PÓS-
GRADUAÇÃO EM CIÊNCIAS AGRÁRIAS - AGRONOMIA

MACHINE LEARNING PARA DETECÇÃO DE *Pratylenchus*
brachyurus NA CULTURA DA SOJA UTILIZANDO
IMAGENS MSI/SENTINEL-2

Autor: Lorena Cristina Alves Nogueira
Orientador: Prof. Dr. Alaerson Maia Geraldine
Coorientador: Prof. Dr. Tavvs Micael Alves

Dissertação apresentada como parte das exigências para obtenção do título de MESTRE EM CIÊNCIAS AGRÁRIAS - AGRONOMIA ao Programa de Pós-Graduação em Ciências Agrárias - Agronomia do Instituto Federal de Educação, Ciência e Tecnologia Goiano – Campus Rio Verde – Área de concentração em Produção Vegetal Sustentável do Cerrado.

RIO VERDE - GO
Novembro - 2020

Sistema desenvolvido pelo ICMC/USP
Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas - Instituto Federal Goiano

NN778m Nogueira, Lorena Cristina Alves Nogueira
MACHINE LEARNING PARA DETECÇÃO DE *Pratylenchus*
brachyurus NA CULTURA DA SOJA UTILIZANDO IMAGENS
MSI/SENTINEL-2 / Lorena Cristina Alves Nogueira
Nogueira; orientador Alaerson Maia Geraldine; co-
orientador Tavvs Micael Alves. -- Rio Verde, 2020.
68 p.

Dissertação (Mestrado em Programa de Pós Graduação -
Ciências Agrárias - Agronomia) -- Instituto Federal
Goiano, Campus Rio Verde, 2020.

1. Agricultura. 2. Fitonematoides. 3.
Sensoriamento Remoto. 4. RPA. 5. Random Forest. I.
Maia Geraldine, Alaerson , orient. II. Micael Alves,
Tavvs, co-orient. III. Título.

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR PRODUÇÕES TÉCNICO-CIENTÍFICAS NO REPOSITÓRIO INSTITUCIONAL DO IF GOIANO

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia Goiano, a disponibilizar gratuitamente o documento no Repositório Institucional do IF Goiano (RIIF Goiano), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IF Goiano.

Identificação da Produção Técnico-Científica

- Tese Artigo Científico
 Dissertação Capítulo de Livro
 Monografia – Especialização Livro
 TCC - Graduação Trabalho Apresentado em Evento
 Produto Técnico e Educacional - Tipo: _____

Nome Completo do Autor: Lorena Cristina Alves Nogueira

Matrícula: 2018202310140075

Título do Trabalho: MACHINE LEARNING PARA DETECÇÃO DE *Pratylenchus brachyurus* NA CULTURA DA SOJA UTILIZANDO IMAGENS MSI/SENTINEL-2

Restrições de Acesso ao Documento

Documento confidencial: Não Sim, justifique: _____

Informe a data que poderá ser disponibilizado no RIIF Goiano: 07/06/2021

O documento está sujeito a registro de patente? Sim Não

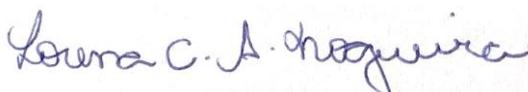
O documento pode vir a ser publicado como livro? Sim Não

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O/A referido/a autor/a declara que:

- o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
- obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia Goiano os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
- cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia Goiano.

Rio Verde, 26 de maio de 2021



Assinatura do Autor e/ou Detentor dos Direitos Autorais

Ciente e de acordo:



Assinatura do(a) orientador(a)



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO

Ata nº 70/2020 - NREPG-RV/CPG-RV/DPGPI-RV/CMPRV/IFGOIANO

ATA Nº/184

BANCA EXAMINADORA DE DEFESA DE DISSERTAÇÃO

Aos trinta dias do mês de novembro do ano de dois mil e vinte, às 09:00h (nove horas), reuniram-se os componentes da Banca Examinadora: Prof. Dr. Alaerson Maia Geraldine (Orientador), Prof. Dr. Tavvs Micael Alves (Avaliador interno) e Prof. Dr. Edipo Henrique Cremon (Avaliador externo), sob a presidência do(a) primeiro(a), em sessão pública realizada por vídeo conferência do IF Goiano – Campus Rio Verde, para procederem a avaliação da defesa de Dissertação, em nível de mestrado, de autoria de **LORENA CRISTINA ALVES NOGUEIRA** discente do Programa de Pós-Graduação em Ciências Agrárias – Agronomia do Instituto Federal Goiano – Campus Rio Verde. A sessão foi aberta pelo presidente da Banca Examinadora Prof. Dr. Alaerson Maia Geraldine, que fez a apresentação formal dos membros da Banca. A palavra, a seguir, foi concedida a (o) autor (a) da Dissertação para, em 40 min., proceder à apresentação de seu trabalho. Terminada a apresentação, cada membro da banca arguiu o (a) examinado (a), tendo-se adotado o sistema de diálogo sequencial. Terminada a fase de arguição, procedeu-se a avaliação da defesa. Tendo-se em vista as normas que regulamentam o Programa de Pós-Graduação em Ciências Agrárias – Agronomia, e procedidas às correções recomendadas, a Dissertação foi APROVADA, considerando-se integralmente cumprido este requisito para fins de obtenção do título de **MESTRE EM CIÊNCIAS AGRÁRIAS-AGRONOMIA**, na área de concentração Produção Vegetal Sustentável no Cerrado, pelo Instituto Federal Goiano – Campus Rio Verde. A conclusão do curso dar-se-á quando da entrega na secretaria do PPGCA-AGRO da versão definitiva da Dissertação, com as devidas correções. Assim sendo, esta ata perderá a validade se não cumprida essa condição, em até **60** (sessenta) dias da sua ocorrência. A Banca Examinadora recomendou a publicação dos artigos científicos oriundos dessa Dissertação em periódicos de circulação nacional e/ou internacional, após procedida as modificações sugeridas. Cumpridas as formalidades da pauta, a presidência da mesa encerrou esta sessão de defesa de Dissertação de Mestrado, e para constar, eu, Vanilda Maria Campos, secretaria do PPGCA-AGRO, lavrei a presente Ata, que, após lida e achada conforme, será assinada pelos membros da Banca Examinadora em duas vias de igual teor.

Prof. Dr. Alaerson Maia Geraldine (Presidente)

Prof. Dr. Tavvs Micael Alves (Membro interno)

Prof. Dr. Edipo Henrique Cremon (Membro externo)

Documento assinado eletronicamente por:

- **Édipo Henrique Cremon**, Édipo Henrique Cremon - Professor Avaliador de Banca - Instituto Federal de Educação, Ciência e Tecnologia de Goiás - Ifg (1), em 30/11/2020 13:21:56.
- **Marconi Batista Teixeira**, COORDENADOR DE CURSO - FUC1 - UCPG-RV, em 30/11/2020 12:50:52.
- **Tavvs Micael Alves**, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 30/11/2020 12:34:44.
- **Alaerson Maia Geraldine**, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 30/11/2020 12:33:02.

Este documento foi emitido pelo SUAP em 26/11/2020. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifgoiano.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 215203

Código de Autenticação: 9cd753cc1b



INSTITUTO FEDERAL GOIANO
Campus Rio Verde
Rodovia Sul Goiana, Km 01, Zona Rural, None, RIO VERDE / GO, CEP 75901-970
(64) 3620-5600

AGRADECIMENTOS

A vida é como uma viagem de trem, as oportunidades são os vagões e as estações e os anos. Agradeço a Deus por estar sempre ao meu lado me guiando, iluminando, dando forças para sempre lutar por todos os vagões e todas as estações que aqui passei, agradeço a ele pela base sólida, pelo porto seguro que me deu, minha família, em especial meus pais, minha irmã e minha avó, por sempre estarem me incentivando, apoiando e cuidando de mim mesmo que de longe, as amizades aqui conquistadas e cultivadas que levarei comigo para sempre.

Ao Programa de Pós-Graduação em Ciências Agrárias – Agronomia do Instituto Federal Goiano – Campus Rio Verde, por abrir as portas principalmente para o meu crescimento profissional e fornecer a infraestrutura necessária para o desenvolvimento das atividades.

Aos Laboratório de Drones e VANT's na Agricultura (LDVA) e Laboratório de Fitopatologia pela disponibilidade dos equipamentos necessários para a pesquisa, a parceria com o professor Dr. Edipo Henrique Cremon do Instituto Federal de Goiás – IFG e a todos os integrantes desta equipe, pois de alguma forma contribuíram para o desenvolver das atividades, as quais não teriam o mesmo sucesso sem a ajuda de cada um.

À CAPES, pela concessão da bolsa de mestrado, uma ajuda financeira inestimável de que todo estudante de pós-graduação necessita.

Por fim, sou meramente grata ao meu orientador Dr. Alaerson Maia Geraldine e Coorientador Dr. Tavvs Micael Alves por terem aceitado esse desafio multidisciplinar junto a mim, pelas orientações, pelos conselhos profissionais quanto pessoais, pela paciência e principalmente por terem acreditado e confiado na minha capacidade, permitindo com que eu fizesse parte dessa equipe.

BIOGRAFIA DO AUTOR

Lorena Cristina Alves Nogueira, nascida na cidade de Patos de Minas – MG, em 28 de fevereiro de 1993 é filha de Lenice Cristina Silva Nogueira e Valdir Alves Nogueira. Iniciou seus estudos em 2011 na Universidade Federal de Uberlândia - Campus Monte Carmelo - MG, onde cursou Engenharia de Agrimensura e Cartográfica. Em 2018 ingressou no Mestrado de Ciências Agrárias - Agronomia pelo Instituto Federal Goiano - Rio Verde – GO, local onde a pesquisa passou a ser vivenciada, praticada e desenvolvida com maior amor e intensidade, com linha de pesquisa em fitopatologia e sensoriamento remoto.

ÍNDICE

	Página
INTRODUÇÃO GERAL	11
1.1 Nematoides de Importância para Agricultura.....	11
1.2. Sensoriamento Remoto para Agricultura	12
1.4. Imagens orbitais do satélite Sentinel-2 MSI.....	13
1.5. Refletância Espectral de Plantas	15
1.6. Índice de Vegetação	16
1.7. Classificação de Imagens Multiespectrais.....	21
1.8. Aprendizagem de Máquina e Reconhecimento de Padrões	22
1.9. Avaliação e Calibração do Modelo de Classificação	24
1.9.1 Matriz de Confusão	24
1.9.2 Índice Kappa	25
1.9.3 Exatidão Global.....	26
Referências Bibliográficas	26
JUSTIFICATIVA.....	34
OBJETIVO GERAL.....	35
Objetivos Específicos.....	35
1.Introdução	38
2. Material e Métodos.....	42
2.1 Área de estudo	42
2.3 Identificação de zonas de amostragem de nematoides.....	42
2.4 Identificação de nematoides no solo e raízes	43
2.5 Aquisição de imagens orbitais	44

2.6	Processamento de imagens multiespectrais e índices de vegetação	45
2.7	Algoritmos de análise e classificação de dados	46
2.8	Treinamento e previsões dos modelos	47
2.9	Importância variável através do índice de Gini	48
2.10	Avaliação da exatidão dos modelos classificados	48
2.11	Produção mapas temáticos para cada modelo	49
3.	Resultados	49
3.1	Parametrização de Random Forest	42
3.2	Importância preditiva das variáveis de entrada	50
3.3	Avaliação de exatidão geral de cada modelo	53
3.4	Comparação entre exatidão e qualidade de índice Kappa de cada modelo gerado.	53
4.	Discussão	55
5.	Conclusões	57
6.	Referências Bibliográficas	58

ÍNDICE DE TABELAS

	Página
Tabela 1. Qualidade da classificação associada aos valores do índice Kappa.	18
Tabela 2. Características das bandas multiespectrais do sensor MSI/Sentinel-2A e 2B	39
Tabela 3. Índices de vegetação calculados a partir da reflectância das imagens multiespectrais do sensor Sentinel-2 MSI conforme base de dados <i>Index Data base</i>	39
Tabela 4- Definição do número de variáveis preditoras (<i>mtry</i>) para cada época de classificação.	41
Tabela 5 – Datas das cenas utilizadas, número de variáveis preditoras, parâmetro de regressão (<i>mtry</i>), exatidão e índice Kappa de modelos de classificação utilizando <i>Random Forest</i>	44

ÍNDICE DE FIGURAS

	Página
Figura 1. Comportamento espectral da vegetação a partir das principais bandas de absorção de clorofila e água. Fonte: Novo, 2010.....	8
Figura 2 – Representação de uma matriz de confusão. Fonte: Adaptado de Congalton and Green, 2008. Em que; n : amostras são distribuídas em K^2 células; n_{ij} : número de amostras classificadas na classe i ($i = 1, 2, \dots, k$) na classificação e na classe j ($j = 1, 2, \dots, k$) no conjunto de dados de referência.....	17
Figura 3. Localização geográfica da área de estudo no município de Rio Verde – GO.....	36
Figura 4. Representação da geometria dos pontos de coleta em campo em cada reboleira de nematoide.....	37
Figura 5 - Variáveis mais importantes para cada modelo classificado por Importância de Gini, considerando a importância dos melhores índices de vegetação e as melhores datas. Época 1 (a), época 2 (b), época 3 (c) e época 4 (d)	45
Figura 6 – Exatidão média entre os modelos classificados pelo algoritmo Random Forest para cada época e interpretação visual através do índice de concordância Kappa.....	47
Figura 7 - Classificação da área de estudo para identificação de ocorrência de nematoides referente as espécies de <i>P. brachyurus</i> , obtida pelo classificador Random Forest utilizando imagens orbitais disponíveis pelo sensor MSI/Sentinel-2A e 2B sugestiva. Época 1 (a), época 2 (b), época 3 (c) e época 4 (d)	47

LISTA DE SÍMBOLOS, SIGLAS, ABREVIACÕES E UNIDADES

Símbolo/Sigla	Significado/Descrição	Unidade de medida
SR	Sensoriamento Remoto	-
SIG	Sistemas de Informação Geográfica	-
RPAs	Remotely Piloted Aircraft	-
DN	Digital number	-
IVs	Índices de vegetação	-
NDVI	<i>Normalized Difference Vegetation Index</i>	-
NDVI ₆₉₀₋₇₁₀	<i>Normalized Difference Vegetation Index 690-710</i>	-
GLI	<i>Green Leaf Index</i>	-
ARVI	<i>Atmospheric resistant vegetation index</i>	-
LCI	<i>Leaf Chlorophyll Index</i>	-
ARVI2	<i>Atmospherically Resistant Vegetation Index 2</i>	-
CHLgreen	<i>Green Atmospherically Resistant Index</i>	-
CARI	<i>Chlorophyll Absorption Ratio Index</i>	-
CARI2	<i>Chlorophyll Absorption Ratio Index 2</i>	-
CCCI	<i>Canopy Chlorophyll Content Index</i>	-
CVI	<i>Chlorophyll vegetation index</i>	-
MCARI	<i>Modified Chlorophyll Absorption in Reflectance Index</i>	-
MTVI2	<i>Modified Triangular Vegetation Index 2</i>	-
RF	Classificador Random Forest	-
mtry	Número de variáveis de entrada	-

ntree	Número de árvores	-
%	Porcentagem	-
Nm	Nanômetros	-
RGB	Comprimento de onda do visível	nm
R	Valor de reflectância no comprimento de onda do vermelho	nm
G	Valor de reflectância na faixa espectral do verde	nm
B	Valor de reflectância na faixa espectral do azul	nm
NIR	Infravermelho Próximo	-
RedEdge	Bordo do Vermelho	-
REM	Radiação eletromagnética	-
IAF	Índice de área foliar	-
°C	Graus celsius	°C
M	Metros	-
ESA	Agência Espacial Europeia	-
GMES	Comissão Europeia de Monitoramento Global do Ambiente e Segurança	-
S2A	Sentinel-2A	-
S2B	Sentinel-2B	-
K	Kappa	-
EG	Exatidão Global	-
Km	Quilômetros	-
R1	Início de Florescimento	-
MP	Mega pixel	-
Cm	Centímetros	-
LDVA	Laboratório de VANT's na agricultura	-
G	Gramas	-
ToA	Reflectância no topo da atmosfera	-
R	Software estatístico	-

RESUMO

NOGUEIRA, LORENA CRISTINA ALVES, Instituto Federal de Educação, Ciência e Tecnologia Goiano-Campus Rio Verde, novembro de 2020. **Machine learning para detecção de *Pratylenchus brachyurus* na cultura da soja utilizando imagens de Sentinel 2.** Orientador: Dr. Alaerson Maia Geraldine. Coorientador: Dr. Tavvs Micael Alves.

A soja é a cultura mais importante do agronegócio brasileiro. Os nematoides podem comprometer sua produtividade e estabilidade econômica mundial. O sensoriamento remoto é uma ciência capaz de proporcionar ferramentas para minimizar os prejuízos causados pelos nematoides. Monitoramento e classificações derivadas de sensores remotos possuem grande potencial para tomada de decisão sobre o manejo de nematoides. Classificadores baseados em aprendizado de máquina são cada vez mais utilizados em alternativas as técnicas tradicionais de classificação de imagens. Este estudo objetivou avaliar o potencial de algoritmo de aprendizado de máquina para classificação temporal de imagens orbitais obtidas pelo sensor multispectral imager (MSI) acoplado ao satélite Sentinel-2. Índices de vegetação foram utilizados para a predição de agrupamentos de plantas infectadas pelo nematoide *Pratylenchus brachyurus* na cultura da soja. Para validação da ocorrência desses fitoparasitas em campo foi realizado um voo exploratório com um drone da área de interesse, para identificar os possíveis agrupamentos de plantas infectadas, foram identificadas cinco zonas com plantas de soja que apresentaram características dos sintomas causados por nematoides, realizou-se coletas de solo e raízes de plantas para análises laboratoriais. No processo de aquisição das imagens orbitais, foram escolhidas imagens durante todo o ciclo da cultura, ou seja, outubro de 2018 a janeiro de 2019. Os dados foram divididos em agrupamentos de imagens para definir quatro modelos de classificação. Para a elaboração dos modelos de classificação utilizou-se o classificador *Random Forest*, a avaliação da classificação foi avaliada por validação cruzada de 10-folds, exatidão global e índice Kappa. Foi calculado uma métrica de importância através do índice de Gini para as variáveis utilizadas nos modelos. O melhor modelo para a detecção de *P. brachyurus* foi o modelo que utilizou todas as variáveis preditoras disponíveis, ou seja, época 4, o índice de vegetação classificado como o mais sensível a identificar agrupamentos de plantas infectadas por nematoides para essa época foi o GLI. Dessa forma, o classificador *Random Forest* apresentou resultados satisfatórios e poderá ser utilizado com alternativa a análise multitemporal para predição da ocorrência de nematoides em cultura de soja.

PALAVRAS-CHAVE: manejo de pragas, floresta aleatória, RPA, fitonematoides, parasitas, agricultura de exatidão.

ABSTRACT

NOGUEIRA, LORENA CRISTINA ALVES, Goiano Federal Institute of Education, Science and Technology - Campus Rio Verde, August 2020. **Machine learning to detect *Pratylenchus brachyurus* in soybean culture using sentinel images 2.** Advisor: Dr. Alaerson Maia Geraldine. Co-advisor: Dr. Tavvs Micael Alves.

Soybean is the most important crop for agribusiness. Nematodes can threaten soybean yield and worldwide economy. Remote sensing is a science with many technological tools used to prevent yield losses by nematodes. Monitoring and remote sensing classifications have potential for better decision-making. Classifiers based on machine learning are increasingly used as alternatives to traditional image classification techniques. This study aimed to evaluate the potential of machine learning algorithms for temporal classification of orbital images obtained by the multispectral imager (MSI) coupled to Sentinel-2 satellite. Vegetation indexes were used to predict clusters of plants infected by the nematode *Pratylenchus brachyurus* in soybean. For validating the occurrence of these phytoparasites in the field, an exploratory flight was conducted using a quadricopter drone to identify possible clusters of infected plants. Five zones showing visible symptoms of nematode-infested plants were identified. Soil and plant roots were collected for laboratory analysis. Orbital images were chosen throughout the crop cycle from October 2018 to January 2029. The data were divided into four classification models using the Random Forest classifier. The classification accuracy was determined by a 10-fold cross validation, global accuracy, and Kappa index. An importance metric was calculated using the gini index for the variables used on models. The best model for detection of *P. brachyurus* used all available predictor variables. The vegetation index GLI was the most sensitive to identify clusters of nematode-infested plants. Thus, the Random Forest classifier presented satisfactory results and can be used as an alternative to multitemporal analysis to predict the occurrence of nematodes in soybean crops.

KEY WORDS: pest management, RPA, phytonematoids, parasites, precision agriculture.

INTRODUÇÃO GERAL

1.1 Nematoides de Importância para Agricultura

Os fitonematoides vivem no solo parasitando plantas. Os fitonematoides podem ser classificados como endoparasitas quando a maior parte de seu ciclo de vida ocorre dentro das raízes das plantas; ou ectoparasitas quando ocorrem externamente a raiz (Forge et al., 1998). Existem mais de 100 espécies e a cerca de 50 gêneros de nematoides espalhados pelo mundo. No Brasil, os principais fitonematoides que atacam a cultura da soja são os nematoides de galhas radiculares *Meloidogyne* spp., nematoide de cisto *Heterodera glycines* e os nematoides das lesões radiculares *Pratylenchus brachyurus* (Gabia, 2019).

Meloidogyne spp. são endoparasitas sedentários que atacam diretamente as raízes da planta. Nas áreas onde ocorrem, observam-se plantas agrupadas em formato irregular com porte pequeno e amareladas (também conhecida como reboleiras), necroses entre as nervuras, desfolha, ausência de raízes e as vezes pode ocorrer abortamento e amadurecimento prematuro de vagens (Hussey et al., 1973; Boneti et al., 1981; Bettiol et al., 2004).

Heterodera glycines é o um dos nematoides mais encontrados na cultura da soja e tem limitado o potencial de expansão e produtividade das áreas plantadas (Ali et al., 2017). Os sintomas observados no campo são reboleiras de plantas com porte reduzido, folhas cloróticas, abortamento de flores e vagens, deterioração das raízes e haste verde ao final do ciclo (Zhang et al., 2017). No entanto, esses sintomas podem ser confundidos com outros problemas, como sintomas causados por deficiência de manganês, infestações de ácaros, danos causados por herbicidas, deficiência de potássio, estresse hídrico e compactação do solo (Godoy et al., 2014). A intensidade dos sintomas depende da

população de nematoides e fertilidade do solo. As condições ideais para emergência dos juvenis e penetração nas raízes das plantas são a alta umidade do solo e temperaturas entre 20 e 30°C. O ciclo de vida completo de *H. glycines* dura a cerca de quatro semanas (Bajwa et al., 2017; Song et al., 2017).

Pratylenchus brachyurus é um endoparasita migrador que vem causando perdas econômicas significativas na cultura de soja, milho, café e pastagem (Dias et al., 2010; Henning et al., 2014). Esse nematoide desenvolve melhor com alta umidade no solo e o seu ciclo completo pode variar de três a seis semanas, sobrevivendo por vários meses dentro de plantas hospedeiras e causando lesões no sistema radicular (Good et al., 1958; Dias-Arieira et al., 2018). Os principais sintomas apresentados pelas plantas infectadas por esses nematoides são: porte reduzido, raízes necrosadas e de coloração escura. Os sintomas podem não ser visualmente evidentes, mas geralmente, as plantas observadas em alguns momentos do desenvolvimento da cultura como aparentemente saudáveis podem estar infestadas com nematoides (Goulart, 2008).

1.2. Sensoriamento Remoto para Agricultura

O processo de detecção e amostragem de nematoides em áreas tradicionalmente requer amostras de solo e raiz, o que dificulta a amostragem em grandes áreas (Martins et al., 2017). O sensoriamento remoto (SR) é uma alternativa para detecção e amostragem direcionada de nematoides com menor custo e maior agilidade. O SR é a ciência amplamente utilizada no estudo da interação da radiação eletromagnética com objetos, sem o contato direto entre o objeto e o sensor (Novo, 2010). Essa técnica envolve a obtenção de dados espectrais por várias plataformas, como satélites artificiais orbitais ou suborbitais, tais como aviões, RPAs (Aeronave Remotamente Pilotada) e sensores terrestres de curta distância como espectrorradiômetros.

Para o estudo dos comportamentos espectrais de um determinado alvo, os equipamentos transformam energia eletromagnética em produtos passíveis de interpretação, geralmente registrado como valores digitais (Novo & Ponzoni, 2001). Esses equipamentos podem ser sensores embarcados em plataformas terrestres, que são classificados como sistema sensor não imageadores, por apresentar informações em formato de gráficos e números e em plataformas suborbital/orbital que são classificados como sistemas sensores imageadores, uma vez que os dados obtidos através da radiação eletromagnética (REM) capturada é observada através de imagens (Novo, 2010; Moreira, 2005; Abreu & Coutinho, 2014).

Os sistemas sensores também podem diferir em resoluções espaciais, espectrais, radiométricas e temporais, as quais irão determinar a qualidade dos produtos gerados. As imagens de sensores remotos possuem estrutura matricial, e seu elemento principal é denominado pixel (Novo, 2010; Ponzoni & Shimabukuro, 2012). De acordo com Novo (2010), os tipos de resoluções são definidos como:

a) A resolução espacial de uma imagem se refere a menor unidade distinguível da imagem que o sensor consegue registrar que representará uma área com as mesmas dimensões na superfície da Terra. Quanto menor a área quadrada do pixel, maior é a resolução espacial da imagem, conseqüentemente, melhor será a definição dos alvos.

b) A resolução espectral está relacionada com a largura da faixa espectral e seu posicionamento ao longo do espectro eletromagnético e o número de bandas, que os sensores conseguem registrar, cada sensor trabalha com um intervalo correspondente destas ondas emitidas. Esta resolução é adotada para a identificação de alvos distintos, indicando a região de origem do alvo no espectro eletromagnético, ou seja, o tipo de onda/banda registrada.

c) A resolução radiométrica é determinada a partir do nível de cinza (cada pixel possui um atributo Z, que indica o DN – digital number) que vai variar do preto ao branco. Essa escala em tons de cinza representa a intensidade média de energia eletromagnética (refletida/emitida) registrada pelo sensor. Cada sensor possui um limite de tons de cinza que consegue detectar e armazenar, mensurado em bits.

d) A resolução temporal é o tempo em dias em que o sensor leva para revistar uma mesma área, geralmente em dias ou horas. Essa informação é de grande importância, principalmente para análises multitemporais, levantamentos e previsão das datas das imagens.

1.4. Imagens orbitais do satélite Sentinel-2 MSI

As imagens de satélite podem ser utilizadas para coleta de informações adquiridos regularmente sobre as culturas em grandes áreas, sendo um sistema confiável e global de monitoramento. Porém, é difícil obter uma série de imagens sem nuvens em alta resolução espacial por um longo período (Thenkabail et al., 2012; Pinter et al., 2003). A Agência Espacial Europeia (ESA) em conjunto com a Comissão Europeia de Monitoramento Global do Ambiente e Segurança (GMES) criou o programa Copernicus com o objetivo de reduzir os intervalos entre imagens de qualidade e aumentar a diversidade de dados (ESA, 2015, 2019).

A missão do Sentinel-2 do programa Copernicus tem o objetivo principal de imageamento multiespectral da superfície terrestre de alta resolução espacial e temporal para monitoramento da vegetação, cobertura do solo e recursos hídricos (Hagolle et al., 2015; ESA, 2015, 2019). A plataforma é constituída por uma constelação de dois satélites gêmeos: o Sentinel-2A (S2A) e o Sentinel-2B (S2B) lançados em 23 de junho de 2015 e 7 de março de 2017, respectivamente. Estes satélites operam simultaneamente com separação de 180°, órbita polar síncrona ao sol, altitude média de 786 km e recobrimento de toda a superfície terrestre entre as faixas de latitude de 56° Sul (Cabo Horn e América do Sul) e 84° Norte (acima da Groenlândia), com uma largura de imageamento de 290 km. A vida útil dos satélites foi estimada em aproximadamente 7,25 anos de operação, com baterias e propelentes capazes de acomodar 12 anos de operação, incluindo as manobras de decaimento ao final da vida útil (Drusch et al., 2012; ESA, 2019;).

O satélite sentinel-2 é composto por instrumentos multiespectrais (MSI) com resoluções espaciais de 10, 20 e 60 metros e resolução temporal de 10 dias que quando combinado com o Sentinel-2B passa então a ser de 5 dias. Os dados possuem resolução radiométrica de 12 bits e opera em largura de 13 canais espectrais com comprimentos de onda do visível (RGB) até o SWIR. O sensor possui algumas bandas estreitas, particularmente nas partes do espectro de borda vermelha e infravermelho próximo (NIR) para avaliação da condição da vegetação além de possui uma largura de faixa de imageamento. O potencial de aplicação destes produtos vem se mostrando cada vez mais vantajoso quando comparado a outros satélites. (ESA, 2019).

Estudos mostraram que dados obtidos através do Sentinel-2 MSI são capazes de estimar a biomassa de pastagens (Sibanda et al., 2015), aplicação de índices de vegetação (Atzberger & Richter, 2012; Hill, 2013), teores de clorofila e nitrogênio foliar (Clevers & Gitelson, 2013; dos Santos et al, 2019), estudos geológicos (Van der Werff & Van der Meer, 2015), mapeamentos de recifes de corais (Hedley et al., 2012), estimativa de parâmetros e variáveis biofísicas (Frampton et al., 2013), estudos arqueológicos (Agapiou et al., 2014), qualidade da água (Salama et al., 2012), mapeamentos de uso e ocupação do solo (Borràs et al., 2017). Esses estudos mostraram que através das características espectrais do Sentinel-2 MSI foi possível obter resultados satisfatórios a partir de vários objetivos envolvendo sensores multiespectrais.

1.5. Refletância Espectral de Plantas

A folha é o principal órgão absorvedor de radiação eletromagnética (REM) das plantas. Dessa forma, as características espectrais de uma folha resultam de sua composição morfológica, química e estrutura interna (Klering et al., 2014; Novo, 2010). As plantas interagem com a REM de forma diferente dependendo do comprimento de onda observado. Nesse sentido, a radiação solar ao atingir o dossel das plantas pode ser parcialmente absorvida pelos pigmentos das folhas para o processo fotossintético. Outra parte da radiação pode ser refletida dependendo dos espaços ocupados por água e ar, pigmentos presentes nas folhas e a estrutura celular interna. Outra parte da REM pode ser transmitida pelas camadas de folhas (Moreira, 2005, Ponzoni et al., 2007).

As plantas se adaptam para absorver apenas o necessário de REM e refletir a radiação em faixas do comprimento de onda que não são fotossinteticamente ativos. Em função dos pigmentos fotossintéticos (clorofila e carotenoides) presentes nas folhas, a radiação incidente é absorvida na região do visível de 400 a 700 nm do espectro eletromagnético (Figura 1), sendo essas regiões espectrais fundamentais para realização da fotossíntese. Enquanto isso, no infravermelho próximo (NIR) compreende a faixa espectral entre 700 e 1300 nm, as folhas exibem altos valores de reflectância e transmissão, principalmente relacionados com as propriedades estruturais e biomassa foliar. A refletância espectral em comprimentos de onda mais longos entre 1300 e 2500 nm é altamente influenciada pela absorção de água pelas plantas (Daughtry et al., 2000; Lillesand & Chipman, 2014).



Figura 1. Comportamento espectral da vegetação a partir das principais bandas de absorção de clorofila e água. Fonte: Novo, 2010.

Estudos mostram que o conteúdo de clorofila desempenha papel direto na fotossíntese e responde a uma variedade de tensões, sendo então frequentemente usado como indicador geral de saúde das plantas e interpretados juntamente com os sinais relacionados à biomassa e ao teor de água. Diante das variações climáticas sazonais, pragas, doenças, deficiência nutricional, propriedades do solo, mudança de estágios fenológicos, as plantas sofrem alterações no seu padrão de reflectância espectral ao longo do seu ciclo (Ponzoni & Disperati, 2001), diante disso, essas mudanças podem ser observadas através de dados capturados por sensores embarcados em satélite. O uso de técnicas de SR para monitorar e detectar essas situações presentes no dia a dia da lavoura vem sendo usada para melhor tomada de decisão.

A detecção de algum estressor na planta por meio de técnicas de SR é fundamentada na radiação refletida pelas folhas. A quantidade e qualidade da REM refletida pela folhagem depende da sanidade do tecido vegetal, que ao ser infectado/contaminado por algum patógeno reage através de algum sintoma. Esses sintomas possuem papel muito importante para detecção espectral, pois, essas plantas infectadas apresentam menor reflectância na região do infravermelho, quando comparadas com tecidos saudáveis, sendo assim possível sua detecção e quantificação pelas diferenças espectrais (Nutter et al., 2002; Naue et al., 2011). Plantas infectadas por nematoides normalmente apresentam crescimento reduzido e sintomas de clorose no dossel, e proporcionam contraste na resposta espectral diferente da reflectância de plantas saudáveis (Dias-Arieira et al., 2010).

1.6. Índices de Vegetação

Os índices de vegetação são combinações aritméticas de bandas espectrais em determinados comprimentos de onda capturadas por sensores. Essas combinações de valores espectrais podem ser adicionadas, divididas ou multiplicadas de maneira que cada pixel da imagem indique a quantidade ou vigor da vegetação. Os índices de vegetação são formas simples e eficientes para análise de vários parâmetros associados a sanidade de vegetação como, por exemplo, em metodologias de realce e identificação de alvos, além de apresentarem a possibilidade de simplificar o processo de mapeamento e classificação de imagens (Campbell & Wynne, 2011; Rey-Caramés et al., 2015).

Os índices de vegetação também podem ser definidos como subproduto do SR. O uso dessa técnica possibilita analisar a distribuição espacial das áreas cultivadas,

melhorando o direcionamento das amostragens e observações de campo (Machado, et al., 2003). O desempenho de cada índice de vegetação depende do que se quer analisar e podem variar ao longo do desenvolvimento da vegetação, ou seja, cada índice é específico para cada característica físico-química de interesse (Poças et al., 2012). Índices de vegetação compostos por bandas no comprimento de onda do visível e do infravermelho possui forte correlação com a biomassa, índice de área foliar (IAF) e conteúdo de clorofila. Conforme o dossel da vegetação sadia se desenvolve, a reflectância aumenta no comprimento de onda do infravermelho próximo e a sua absorção aumenta na região do vermelho pelo aumento de fotossíntese (Jensen, 2009; Ponzoni et al., 2012).

Existem diversos índices de vegetação desenvolvidos para avaliação de parâmetros da agricultura, e alguns desses pode ser calculados através dos dados obtidos pelo sensor MSI/Sentinel-2A e 2B, tais como: NDVI - *Normalized Difference Vegetation Index*; NDVI₆₉₀₋₇₁₀ - *Normalized Difference Vegetation Index 690-710*; GLI - *Green Leaf Index*; ARVI - *Atmospheric resistant vegetation index*; LCI - *Leaf Chlorophyll Index*; ARVI2 - *Atmospherically Resistant Vegetation Index 2*; CHLgreen - *Green Atmospherically Resistant Index*; CARI - *Chlorophyll Absorption Ratio Index*; CARI2 - *Chlorophyll Absorption Ratio Index 2*; CCCI - *Canopy Chlorophyll Content Index*; CVI - *Chlorophyll vegetation index*; MCARI/MTVI2 - *Modified Chlorophyll Absorption in Reflectance Index / Modified Triangular Vegetation Index 2*.

O índice de vegetação por diferença normalizada (NDVI) aplica a diferença da reflectância entre a faixa do infravermelho próximo (NIR) e a reflectância da faixa de vermelho (Eq. 1) (Duan et al., 2017). A vegetação saudável absorve a maior parte da luz dos comprimentos de onda do visível (RGB) e reflete grande parte na banda de infravermelho. Quando a vegetação está doente, a planta reflete maior quantidade de luz nos comprimentos de onda do visível (RGB) e menor quantidade no infravermelho próximo, desse modo, é possível obter avaliações qualitativas e quantitativas da vegetação (Riedell & Blackmer, 1999; Zhang et al. 2003; Marin et al., 2019). Embora seja considerado um índice com bons resultados para análise da sanidade da vegetação, para sua interpretação deve-se considerar a saturação do pixel em áreas com vegetação densa, a geometria entre sol-alvo-sensor e angulação solar deve ser consideradas para a interpretação (Huete et al., 1999; Ponzoni et al., 2007; Jensen, 2009).

$$NDVI = \frac{(NIR - R)}{(NIR + R)} \quad (1)$$

Em que,

R: valor de reflectância no comprimento de onda do vermelho;

NIR: valor de reflectância no comprimento de onda do infravermelho próximo.

O índice de vegetação por diferença normalizada - NDVI₆₉₀₋₇₁₀ foi desenvolvido para fornecer uma estimativa precisa do teor de clorofila foliar utilizando o comprimento de onda NIR com comprimentos de ondas específicos de 690-710 nm (Eq. 2), estudos mostraram que a refletância próxima ao comprimento de onda de 710 nm apresentou maior sensibilidade ao conteúdo de clorofila do que a refletância próxima a 550 nm, pois os efeitos do espalhamento foliar na refletância está fortemente relacionado com a variação na refletância causada pela absorção da clorofila (Datt, 1999 & Ponzoni et al., 2012).

$$\text{NDVI}_{690-710} = \frac{\text{NIR} - [690:710]}{\text{NIR} + [690:710]} \quad (2)$$

Em que,

NIR: valor de reflectância no comprimento de onda do infravermelho próximo;

690:710: valor de reflectância no comprimento de onda de 690-710 nm.

O índice de vegetação de folha verde – GLI é formado pela diferença normalizada entre os valores de reflectância obtidos pelas faixas espectrais do visível (RGB) (Eq. 3). O índice GLI tem como objetivo determinar a intensidade de degradação da vegetação por fotografia aérea. Louhaichi, Borman & Johnson (2001) acreditam que a utilização deste índice tem potencial promissor de aplicação para análises vegetais, uma vez que apresentou resultados satisfatórios para avaliação da fração da vegetação. Alguns autores mostraram bons resultados quando utilizado o índice GLI para avaliação da vegetação apresentado boas correlações com alterações na clorofila foliar, com objetivo de avaliar o fator indicativo de degradação vegetal (Hunt JR. et al., 2011; Ballesteros et al., 2018).

$$\text{GLI} = \frac{(G - R) + (G - B)}{(2 * G) + R + B} \quad (3)$$

Em que,

R: valor de reflectância na faixa espectral do vermelho;

G: valor de reflectância na faixa espectral do verde;

B: valor de reflectância na faixa espectral do azul.

O índice de vegetação atmosféricamente resistente – ARVI é relativamente propenso a fatores atmosféricos. Esse índice foi desenvolvido por Kaufman & Tanré, 1992 a partir do índice NDVI, que tinha como objetivo, criar uma fórmula (Eq. 4) em que

apresentasse menor sensibilidade a efeitos atmosféricos, como efeitos de espalhamento e aerossol, e por meio de um processo de autocorreção dos efeitos da atmosfera no comprimento de onda do vermelho encontrasse melhores resultados. Além de ser um índice bastante utilizado na agricultura, o ARVI também é mais robusto aos efeitos topográficos, e o tornar uma ferramenta de monitoramento altamente eficaz para regiões montanhosas tropicais frequentemente poluídas com fumaças proveniente da agricultura de corte e queimadas (Gitelson et al., 2002; Mokarram et al., 2016, Mckinnon & Hoff, 2017).

$$ARVI = \frac{(NIR - (2 * R) + B)}{(NIR + (2 * R) + B)} \quad (4)$$

Em que,

NIR: valor de reflectância no comprimento de onda do infravermelho próximo;

R: valor de reflectância na faixa espectral do vermelho;

B: valor de reflectância na faixa espectral do azul.

O índice de proporção de absorção de clorofila – CARI foi desenvolvido por Kim et al., (1994) (Eq. 5) com objetivo de reduzir fatores não fotossintéticos através da medida da profundidade da absorção de clorofila a 670 nm em relação ao pico de reflectância verde a 550 nm e a reflectância a 700 nm.

$$CARI = \left(\frac{700nm}{670nm} \right) * \left(\frac{\sqrt{(a - 670 + 670nm + b)^2}}{(a^2 + 1)^{0.5}} \right) \quad (5)$$

Em que,

550; 670; 700: valor de reflectância no comprimento de onda de 550, 670 e 700 nm;

$a = (700nm - 550nm) / 150$;

$b = (550nm - ((700nm - 550nm) / 150 * 550))$.

O índice de clorofila foliar - LCI é aplicado para calcular o teor total de clorofila das folhas e é menos afetado pelas variações da estrutura interna e pelo espalhamento da superfície foliar. O LCI é uma razão entre as profundidades de absorção relativa da clorofila nos comprimentos de onda de 710 e 680 nm (Eq. 6). Funciona melhor em níveis moderados a altos de clorofila. O valor de LCI aumenta à medida que o teor de clorofila nas folhas aumentam (Datt, 1999).

$$LCI = \frac{[850] - [710]}{[850] + [680]} \quad (6)$$

Em que,

680 nm: comprimento de onda específico compreendido na região do vermelho;
710 e 850 nm: comprimento de onda específico compreendido na região do infravermelho.

O índice de absorção de clorofila modificada no índice de refletância / índice de vegetação triangular modificado 2 (MCARI/MTVI2; Eq. 7) é mais sensível às variações do teor de clorofila e mais resistente às variações de Índice de área foliar (IAF) e ângulo do zênite solar do que um único índice de vegetação (Haboudane et al., 2002). Para melhorar a sensibilidade do índice de vegetação a variações sutis na concentração de clorofila nas folhas de arroz sob estresse por metais pesados (Wu et al., 2008), indicaram que a integração entre índices apresenta melhor linearidade com o teor de clorofila.

O índice MCARI (*Modified Chlorophyll Absorption in Reflectance Index*) foi arquitetado visando medir a clorofila nos dosséis agrícolas. O MCARI fornece uma medida da profundidade da absorção da clorofila e é muito sensível a variações nas concentrações de clorofila. O MCARI utiliza bandas correspondentes à absorção mínima dos pigmentos fotossintéticos, compreendida em 550 e 700 nm, em conjunto com a clorofila, uma banda de absorção máxima de 670 nm. Os resultados obtidos pelo índice não são afetados pelas condições de iluminação, pela refletância de fundo do solo e outros materiais não fotossintéticos (Kim et al., 1994; Daughtry et al., 2000). Os modelos empíricos para a previsão do teor de clorofila a partir da refletância comumente são baseados na refletância a cerca de 550 ou 700 nm, e a saturação do efeito de absorção ocorre apenas com um teor mais alto de clorofila (Sims & Gamon, 2002; Wu et al., 2008).

O índice MTVI2 (*Modified Triangular Vegetation Index 2*) foi reformulado com objetivo de ser menos sensível aos efeitos da clorofila, e mais resistentes aos efeitos do solo e da atmosfera (Haboudane et al., 2004). Resultados obtidos através do MTVI2 mostram um comportamento semelhante ao MSAVI, porém, é positivamente menos sensível às mudanças na concentração de clorofila. Alguns autores mostraram que o índice de razão de absorção de clorofila modificado e o segundo índice de vegetação triangular modificado em razão (MCARI / MTVI2) são resistentes ao índice de área foliar (IAF), apresentando baixa correlação com IAF e maior correlação com clorofila, assim puderam concluir que o uso de MCARI / MTVI2 pode melhorar as estimativas no solo do status da cultura em que a variabilidade está associada com a disponibilidade de água (JOUR et al.; 2008).

$$\frac{\text{MCARI}}{\text{MTVI2}} = \frac{((700\text{nm}-670)-0.2(700\text{nm}-550\text{nm}))\left(\frac{700\text{nm}}{670\text{nm}}\right)}{\left(1.5 \frac{1.2(800\text{nm}-550\text{nm})-2.5(670\text{nm}-550\text{nm})}{\sqrt{(2800\text{nm}+1)^2-6800\text{nm}-5\sqrt{670\text{nm}}-0.5}}\right)} \quad (7)$$

1.7. Classificação de Imagens Multiespectrais

Classificar uma imagem consiste em extrair informações espectrais, delimitar e determinar objetos, coisas, espaços, matéria, separando-os em grupos semelhantes/homogêneos (Camara et al., 1996). Os principais critérios que devem ser levados em consideração para o uso de técnicas de classificação de imagens multiespectrais são a seleção das imagens, pré-processamento, seleção do algoritmo de classificação, seleção dos dados de treinamento, definição das assinaturas espectrais das classes, bem como a apresentação do mapa temático (Campbell & Wynne, 2011).

Na metodologia de classificação de imagens pode ser dividida em supervisionada e não supervisionada. Em uma classificação supervisionada são determinadas diferentes amostras de treinamento com identidade/características conhecida, para assim classificar pixels de identidade/características desconhecida (Campbell & Wynne 2011). Essa classificação utiliza as propriedades espectrais de cada pixel das amostras de treinamento como atributo para classificar os demais pixels, assim cada pixel é caracterizado em um tipo de classe já determinada pelo classificador. Na classificação não supervisionada, o usuário não controla a separação das classes de interesse pelo algoritmo (Moreira, 2005; Li et al., 2015).

Modelos matemáticos denominados de algoritmo são utilizados para realizar classificações. Os algoritmos podem ser classificados em paramétricos e não paramétricos. Os paramétricos exigem que as amostras tenham distribuição normal e que as classes possam ser modeladas por distribuições de probabilidade sendo representados pelos parâmetros dessas distribuições, em contrapartida os não paramétricos não necessitam de dados com distribuição normal, não havendo proximidade com os modelos de distribuição de uma curva normal (Schowengerdt, 2006; Rodrigues et al., 2015).

Para os classificadores paramétricos aplicados ao sensoriamento remoto, o classificador Máxima Verossimilhança (MAXVER) é uma das principais soluções com forma pré-definida, ou seja, os dados (pixels) de treinamento se aproximam de uma distribuição normal. Por outro lado, os algoritmos não paramétricos, normalmente baseados em aprendizado de máquina, são bastante promissores para a classificação de

imagens, pois não fazem pré-suposições sobre os dados, permitem viabilizar maior separação entre as classes geradas, possibilitam a inclusão de dados não espectrais, são capazes de extrair informações a partir de amostras/dados de treinamento e em seguida redistribuí-los por toda a imagem (Richards & Jia, 2006; Lary et al., 2016).

1.8. Aprendizagem de Máquina e Reconhecimento de Padrões

Aprendizagem de máquina é uma área da inteligência artificial que estuda métodos computacionais para obter padrões semiautomáticos ou automaticamente, em função de informações existentes (Witten & Frank 2005). O método de classificação de imagens utilizando algoritmos de aprendizagem de máquina consiste na determinação de regras e na classificação dos dados. Este conjunto de regras é criado por um classificador, que recebe como entrada um vetor de características e oferece como saída uma classe resultante para a instância que as características descrevem (Breiman, 2001).

Os algoritmos de aprendizagem de máquina mais utilizados são: Rede Neural Artificial, Máquina de Vetores de Suporte (*Support Vector Machine*) e Floresta Aleatória (Random Forest). Esses podem ser classificados de acordo com o resultado encontrado em cada ambiente como: aprendizagem não supervisionada, aprendizagem supervisionada e aprendizagem por reforço (Russel & Norvig, 2013).

Alguns estudos mostram a eficiência desses algoritmos como técnica de classificação. O classificador Floresta Aleatória foi utilizado com o objetivo de analisar e mapear mudanças na cobertura da terra usando diferentes tipos de dados, como imagens Landsat, temperatura da superfície do solo, modelo digital de terreno (MDT) e textura (Rodriguez-Galiano & Chica-Olmo, 2012). Taghizadeh-Mehrjardi et al., (2015) testaram o potencial de 6 classificadores de aprendizagem de máquina (redes neurais artificiais, máquina de vetor de suporte, K vizinhos mais próximos [*K-nearest neighbour*], floresta aleatória, árvore de decisão e regressão logística) e dados auxiliares como índices espectrais e bandas específicas derivadas do satélite Landsat 8. Esses autores encontraram que os algoritmos de redes neurais artificiais e árvores de decisão foram os algoritmos que apresentaram melhor desempenho para prever as famílias de solos. Através dos métodos de classificação baseados em pixel, redes neurais artificiais, classificador de máxima verossimilhança, classificação baseada em objeto e em combinação com o algoritmo de árvore de decisão foi possível detectar as mudanças no uso e cobertura da terra a partir de imagens multiespectrais em diferentes anos (Landsat MMS, TM, ETM+ e OLI/TIRS) (Minaei & Kainz, 2016). Através da classificação realizada pelo

classificador Multilayer Perceptron. Tran et al., (2017) verificou a relação entre as mudanças no uso e cobertura da terra e temperatura de superfície, além de dados obtidos pelo satélite Landsat 5 TM, Landsat ETM+ e Landsat 8 OLI, também foi utilizado os índices de vegetação NDVI e NDBI. De Alban et al., (2018) testaram o classificador floresta aleatória para detecção de mudanças no uso e cobertura da terra em paisagens tropicais dinâmicas, através de imagens Landsat-5 TM, Landsat-8 OLI, JERS-1 SAR, ALOS/PALSAR-2, índices espectrais e medidas de textura. Utilizando ainda desse mesmo classificador, Chemura et al., (2017) avaliaram o potencial das bandas do Sentinel-2 para detectar os níveis de infecção da ferrugem no café através da refletância das folhas, precocemente devido a suas taxas devastadoras e para análises estatísticas utilizaram Análise Parcial por Mínimos Quadrados (PLS-DA).

O algoritmo Floresta Aleatória é um método de classificação baseado em aprendizagem supervisionada com árvores de classificação e regressão se tratando assim, de um conjunto de classificadores de árvore de decisão. Na fase de treinamento, um número determinado de árvores é gerado utilizando um subconjunto pseudoaleatório do vetor de características completo utilizado no problema. Tanto o número de árvores quanto a quantidade de atributos utilizados por cada árvore pode ser determinado pelo usuário do algoritmo. Uma vez que a base de treinamento foi utilizada para gerar os modelos das árvores de decisão, estes modelos podem ser utilizados para classificar novas amostras do problema. Preferencialmente, todas as árvores geradas são envolvidas na classificação, cada uma chegando a sua própria conclusão sobre a classe da amostra apresentada. Cada árvore tem uma “opinião/resposta”, que é determinado para a definição da classe mais votada, que é então escolhida como a classe da amostra (Breiman, 2001).

A principal vantagem do uso de Floresta Aleatória é a eliminação de *overfitting*, problema bastante comum quando se utiliza árvores de decisão. Breiman, (2001) afirma que a taxa de erro de um modelo de aprendizado criado com Floresta Aleatória está relacionada a dois fatores: a correlação entre quaisquer duas árvores geradas no modelo e a “força” de cada árvore gerada, ou seja, o quão precisa cada árvore é em relação ao modelo geral. À medida que a correlação entre árvores cresce, também cresce a taxa de erro do modelo final. Quanto menor for a taxa de erro de uma árvore individual, mais “forte” é considerado o classificador, e por consequência, menor é a taxa de erro do modelo. Reduzir o número de variáveis do vetor de características a ser utilizado em cada árvore reduz a correlação entre árvores e reduz a “força” da árvore. Aumentar este número de variáveis aumenta a correlação e a “força” da árvore. A parametrização do algoritmo

deve se preocupar em encontrar o número de variáveis do vetor de características que melhor minimize a correlação e maximize a “força” das árvores do modelo.

1.9. Avaliação e Calibração do Modelo de Classificação

Todo modelo criado a partir de uma técnica de aprendizagem de máquina deve ser testado e avaliado com a finalidade de verificar sua adequação e se os atributos escolhidos são importantes para a classificação das amostras. Um parâmetro importante para essa verificação é baseado no quão foi exato a classificação das amostras, ou seja, determinar a acurácia ou taxa de reconhecimento dos resultados obtidos (Gao, 2008).

Existem vários fatores que podem interferir na qualidade da classificação de imagens. Os principais fatores que levam a ocorrência de erros nas classificações visuais estão relacionados com identificação de classes, excessiva generalização, erros de registro e variação nos detalhes de interpretação. Para as classificações digitais os principais erros são referentes à similaridade espectral da vegetação, dos tamanhos das amostras coletadas, número de classes de interesse e número de amostras por classes. A abordagem mais comum para avaliar o resultado de uma classificação é comparando a cena classificada com dados de referência espacialmente e temporalmente coincidentes, os quais são considerados como de maior acurácia (Congalton, 1991; Campbell & Wynne, 2011).

1.9.1 Matriz de Confusão

Um dos métodos mais utilizados para verificar a acurácia da classificação é por meio da matriz de erro denominada de matriz de confusão (Fig. 2). A matriz de confusão é uma matriz quadrada contendo números arranjados por meio de linhas e colunas, que representam o número de pixels atribuídos a uma categoria específica em relação a verdadeira categoria a qual o pixel pertence. As linhas da matriz confusão indicam as classes definidas a partir dos dados de sensoriamento remoto, enquanto as colunas representam os dados de referência. Com base nisso, tem-se que os valores da diagonal principal representam os pixels que foram corretamente classificados (Congalton & Green, 2008).

		j= Coluna (Referência)			Total linha
		1	2	k	
i= Linhas (Classificação)	1	n_{11}	n_{12}	n_{1k}	n_{1+}
	2	n_{21}	n_{22}	n_{2k}	n_{2+}
	K	n_{k1}	n_{k2}	n_{kk}	n_{k+}
Total coluna n_{+j}		n_{+1}	n_{+2}	n_{+k}	n

Figura 2 – Representação de uma matriz de confusão. Fonte: Adaptado de Congalton & Green, 2008. Em que; n : amostras são distribuídas em K^2 células; n_{ij} : número de amostras classificadas na classe i ($i = 1, 2, \dots, k$) na classificação e na classe j ($j = 1, 2, \dots, k$) no conjunto de dados de referência.

1.9.2 Índice Kappa

Algumas medidas descritivas podem ser obtidas por meio da matriz de confusão, tais como: estimativa Kappa (Eq. 8) e acurácia global (Eq. 9). O Índice Kappa é uma medida de concordância que fornece uma ideia do quanto as observações se afastam daquelas esperadas, indicando o quão verdadeiro são as interpretações (Cohen, 1960). O cálculo do índice Kappa é expresso por:

$$k = \frac{[n * \sum_{i=1}^r x_{ij} - \sum_{i=1}^r (x_i * x_j)]}{[n^2 - \sum_{i=1}^r (x_i * x_j)]} \quad (8)$$

Em que;

K: Índice de exatidão Kappa;

r: número de linhas na matriz, em que (i) corresponde ao número de observações na linha e (j) coluna;

n: número total de observações.

A partir dos resultados encontrados, esses são associados aos valores do Índice Kappa à qualidade da classificação de acordo com a tabela abaixo (Tabela 1; adaptado de Landis & Koch, 1977).

Tabela 1. Qualidade da classificação associada aos valores do índice Kappa.

Índice Kappa	Qualidade
0.00	Péssima
0,01 a 0,02	Ruim
0,21 a 0,40	Razoável
0,41 a 0,60	Boa
0,61 a 0,80	Muito boa
0,81 a 1,00	Excelente

Fonte: Adaptado de Landis & Koch, 1977.

1.9.3 Exatidão Global

A Exatidão Global é a medida de verificação que relaciona os elementos da diagonal da matriz confusão com o total de pontos amostrados da imagem classificada, é calculada dividindo a soma dos valores da diagonal principal da matriz, que representa o número de elementos classificados corretamente. O índice global superestima a classificação e sempre resultará em valores mais altos do que os outros índices (Figueiredo et al., 2007; Congalton & Green, 2008).

$$EG = \frac{A}{n} * 100 \quad (9)$$

Em que;

EG: Exatidão Global;

A: número de pontos amostrais com acerto;

n: número de pontos amostrais.

Referências Bibliográficas

- Abreu, K.M.P. de, Coutinho, L.M., 2014. Remote sensing applied to the study of vegetation with emphasis on index of vegetation and landscapes metrics. *Revista Vértices* 16, 173–198.
- Agapiou, A., Alexakis, D.D., Sarris, A., Hadjimitsis, D.G., 2014. Evaluating the potentials of sentinel-2 for archaeological perspective. *Remote Sensing* 6, 2176–2194.
- Ali, M.A., Abbas, A., Azeem, F., Javed, N., Bohlmann, H., 2015. Plant-nematode interactions: From genomics to metabolomics. *International Journal of Agriculture and Biology* 17, 1071–1082.
- Ali, M.A., Azeem, F., Abbas, A., Joyia, F.A., Li, H., Dababat, A.A., 2017. Transgenic

- strategies for enhancement of nematode resistance in plants. *Frontiers in Plant Science* 8, 1–13.
- Atzberger, C., Richter, K., 2012. Spatially constrained inversion of radiative transfer models for improved LAI mapping from future Sentinel-2 imagery. *Remote Sensing of Environment* 120, 208–218.
- Bajwa, S.G., Rupe, J.C., Mason, J., 2017. Soybean disease monitoring with leaf reflectance. *Remote Sensing* 9.
- Ballesteros, R., Ortega, J. F., Hernandez, D., Del Campo, A., Moreno, M. A., 2018. Combined use of agro-climatic and very high-resolution remote sensing information for crop monitoring. *International Journal of Applied Earth Observation and Geoinformation*, 72, 66–75.
- Borràs, J., Delegido, J., Pezzola, A., Pereira, M., Morassi, G., Camps-Valls, G., 2017. Clasificación de usos del suelo a partir de imágenes sentinel-2. *Revista de Teledetección* 2017, 55–66.
- Breiman, L., 2001. Random forests. *Machine learning*, 45, 5-32.
- Camara, G., Cartaxo, R., Souza, M., Freitas, U.M., Garrido, J., 1996. Spring : Integrating Remote Sensing and Gis By Data Modelling. *Science* 20, 395–403.
- Campbell, J. B., and Wynne, R. H., 2011. *Introduction to remote sensing*. Guilford Press.
- Chemura, A., Mutanga, O., Dube, T., 2017. Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions. *Precision Agriculture* 18, 859–881.
- Clevers, J.G.P.W., Gitelson, A.A., 2013. 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* 23, 344–351.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20, 37-46.
- Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37, 35–46.
- Congalton, R.G, Green, K., 2019. *Assessing the accuracy of remote sensing data: principles and practices*. CRC press.
- Crusiol, L., Nanni, M., Furlanetto, R., Silva, G., Silva, M., Sibaldelli, R., Farias, J., 2017. Sensoriamento remoto aplicado ao monitoramento do déficit hídrico na cultura da soja. In *Embrapa Soja- Artigo em anais de congresso (ALICE)*. In: *Jornada Acadêmica da Embrapa soja*, 12, 119-127.
- Datt, B., 1999. A new reflectance index for remote sensing of chlorophyll content in higher plants: Tests using Eucalyptus leaves. *Journal of Plant Physiology* 154, 30–

36.

- Daughtry, C. S. T., Walthall, C. L., Kim, M. S., De Colstoun, E. B., McMurtrey Iii, J. E., 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote sensing of Environment*, 74, 229-239.
- De Alban, J.D.T., Connette, G.M., Oswald, P., Webb, E.L., 2018. Combined Landsat and L-band SAR data improves land cover classification and change detection in dynamic tropical landscapes. *Remote Sensing*, 10, 306.
- Dias-Arieira, C. R., Molina, R. D. O., Costa, A. T., 2010. Nematoides causadores de doenças em frutíferas. *Revista agro Ambiente on-line*, 2, 46-56.
- Dias-Arieira, C. R., de Araújo, F. G., Kaneko, L., Santiago, D. C., 2018. Biological control of *Pratylenchus brachyurus* in soya bean crops. *Journal of Phytopathology*, 166, 722-728.
- Dos Santos, E. P., de Jesus, T. B., Carneiro, A. D. S. R., Santos, R. L., & dos Santos, R. A., 2019. Imagens Sentinel-2A e espectrorradiometria aplicada a estudo de pastagens degradadas. *Anuário do Instituto de Geociências*, 41, 141-151.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment* 120, 25–36.
- Duan, T., Chapman, S.C., Guo, Y., Zheng, B., 2017. Dynamic monitoring of NDVI in wheat agronomy and breeding trials using an unmanned aerial vehicle. *Field Crops Research* 210, 71–80.
- ESA - European Space Agency. 2015. SENTINEL-2 Data access and management. Sentinel Online Handbook 1–2.
<https://sentinel.esa.int/documents/247904/1848117/Sentinel-2_Data_Products_and_Access> (Acessado em janeiro 2020).
- ESA- European Space Agency. 2019. Introducing Sentinel-2.
<<https://sentinel.esa.int/web/sentinel/missions/sentinel-2>>. (Acessado em dezembro de 2019).
- Figueiredo, G. C., Vieira, C. A. O., 2007. Estudo do comportamento dos índices de Exatidão Global, Kappa e Tau, comumente usados para avaliar a classificação de imagens do sensoriamento remoto. *Simpósio Brasileiro de Sensoriamento Remoto*, 13, 5755-5762.
- Forge, T.A., DeYoung, R., Vrain, T.C., 1998. Temporal changes in the vertical distribution of *Pratylenchus penetrans* under raspberry. *Journal of Nematology* 30, 179–183.
- Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing* 82, 83–92.

- Gabia, A. A., 2019. Nematoides de importância econômica na cultura da soja. <<https://promip.agr.br/nematoides-de-importancia-economica-na-cultura-da-soja/>> (Acesso em janeiro 2020).
- Gao, J., 2008. Digital analysis of remotely sensed imagery. McGraw-Hill Professional.
- Gitelson, A. A., Kaufman, Y. J., Stark, R., Rundquist, D., 2002. Novel algorithms for remote estimation of vegetation fraction. *Remote sensing of Environment*, 80, 76-87.
- Godoy, C.V., Manuel, Á., Almeida, R., Soares, R.M., Seixas, D.S., Dias, W.P., Meyer, M.C., Maria, L., 2014. Doenças Da Soja (*Glycine max* (L .) Merrill). Sociedade Brasileira De Fitopatologia (Sbf) 1–14.
- Good, J. M., Boyle, L. W., Hammons, R. O., 1958. Studies of *Pratylenchus brachyurus* on peanuts. *Phytopathology*, 48, 530-535.
- Goulart, A. M. C., 2008. Aspectos gerais sobre nematoides das lesões radiculares (gênero *Pratylenchus*). Embrapa Cerrados-Documents (INFOTECA-E).
- Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J., Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment* 81, 416–426.
- Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., Strachan, I.B., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment* 90, 337–352.
- Hagolle, O., Sylvander, S., Huc, M., Claverie, M., Clesse, D., Dechoz, C., Lonjou, V., Poulain, V., 2015. SPOT-4 (Take 5): Simulation of sentinel-2 time series on 45 large sites. *Remote Sensing* 7, 12242–12264.
- Hedley, J., Roelfsema, C., Koetz, B., Phinn, S., 2012. Capability of the Sentinel 2 mission for tropical coral reef mapping and coral bleaching detection. *Remote Sensing of Environment* 120, 145–155.
- Henning, A. A., Almeida, Á. M. R., Godoy, C. V., Seixas, C. D. S., Yorinori, J. T., Costamilan, L. M., ... Dias, W. P., 2014. Manual de identificação de doenças de soja. Embrapa Soja-Documents (INFOTECA-E).
- Hill, M.J., 2013. Vegetation index suites as indicators of vegetation state in grassland and savanna: An analysis with simulated SENTINEL 2 data for a North American transect. *Remote Sensing of Environment* 137, 94–111.
- Huete, A., Justice, C., Van Leeuwen, W., 1999. MODIS vegetation index (MOD13). Algorithm theoretical basis document, 3, 01-213.

- Hunt Jr, E.R., Daughtry, C.S.T., Eitel, J.U., Long, D. S., 2011 Remote sensing leaf chlorophyll content. *Agronomy Journal*, 103, 1090–1099.
- Hussey, R. S., 1973. A comparison of methods of collecting inocula of *Meloidogyne* spp., including a new technique. *Plant Dis. Rep.*, 57, 1025-1028.
- Jensen, J. R., 2009. Remote sensing of the environment: An earth resource perspective 2/e. Pearson Education India. São José dos Campos, SP: Parêntese, 01-598.
- Joalland, S., Screpanti, C., Varella, H.V., Reuther, M., Schwind, M., Lang, C., Walter, A., Liebisch, F., 2018. Aerial and ground based sensing of tolerance to beet cyst nematode in sugar beet. *Remote Sensing* 10, 01–21.
- Kaufman, Y.J., Tanre, D., 1992. Índice de vegetação resistente à atmosfera (ARVI) para EOS-MODIS. *Transações IEEE em Geociências e Sensoriamento Remoto*, 30, 261-270.
- Kim, M. S., Daughtry, C. S. T., Chappelle, E. W., McMurtrey, J. E., & Walthall, C. L., 1994. The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation (A par).
- Klering, E.V., Fontana, D.C., Alves, R., Rocha, J., Berlato, M.A., 2014. Estimativa De Área Cultivada Com Arroz Irrigado Para O Estado Do Rio Grande Do Sul a Partir De Imagens Modis. *Ciência e Natura* 35, 126–135.
- Landis, J. R., Koch, G. G., 1977. The measurement of observer agreement for categorical data. *Biometrics*, 159-174.
- Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. *Geoscience Frontiers* 7, 3–10.
- Li, X., Lee, W.S., Li, M., Ehsani, R., Mishra, A.R., Yang, C., Mangan, R.L., 2015. Feasibility study on Huanglongbing (citrus greening) detection based on WorldView-2 satellite imagery. *Biosystems Engineering* 132, 28–38.
- Lillesand, T., Kiefer, R.W, Chipman, J. 2015. Sensoriamento remoto e interpretação de imagens. John Wiley & Sons.
- Louhaichi, M., Borman, M. M., & Johnson, D. E., 2001. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto International*, 16, 65-70.
- MacDonald, S.L., Staid, M., Staid, M., Cooper, M.L., 2016. Remote hyperspectral imaging of grapevine leafroll-associated virus 3 in cabernet sauvignon vineyards. *Computers and Electronics in Agriculture* 130, 109–117.
- Machado, A., 2015. Nematoides devastam lavouras de soja. <http://www.revistacampoenegocios.com.br/nematoides-devastam-lavouras-de-soja/>. (Acessado em janeiro de 2020).

- Machado, H. M., Lamparelli, R. A. C., Rocha, J. V., JUNIOR, J. Z., 2003. Estudo da variação de parâmetros atmosféricos necessários para correção atmosférica de imagens de satélite. Simpósio Brasileiro de Sensoriamento Remoto, XI, 2537-2539.
- Marin, D.B., Alves, M. de C., Pozza, E.A., Gandia, R.M., Cortez, M.L.J., Mattioli, M.C., 2019. Multispectral remote sensing in the identification and mapping of biotic and abiotic coffee tree variables. *Revista Ceres* 66, 142–153.
- Martins e Souza Filho, P.W., Paradella, W.R., Souza Júnior, C., Valeriano, D. de M., Miranda, F.P. de, 2006. Sensoriamento remoto e recursos naturais da Amazônia. *Ciênc. cult. (São Paulo)* 58, 37–41.
- Martins, G.D, Galo, M.D.L.B.T, Vieira, B.S. 2017 Detecção e mapeamento da infecção do nematóide das galhas na cultura do café usando medições de sensoriamento remoto. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10, 5395-5403.
- Martins, G.D., Trindade Galo, M.D.L.B., 2015. In situ spectral characterization of sugar cane infected by nematodes and *Migdolus fryanus*. *Boletim de Ciencias Geodesicas* 21, 783–796.
- McKinnon, T., Hoff, P., 2017. Comparing RGB-based vegetation indices with NDVI for drone based agricultural sensing. *Agrirobotix. Com*, 1-8.
- Minaei, M., Kainz, W., 2016. Watershed land cover/land use mapping using remote sensing and data mining in Gorganrood, Iran. *ISPRS International Journal of Geo-Information* 5.
- Mokarram, M., Bolorani, A.D., Hojati, M., 2016. Relationship between land cover and vegetation indices. Case study: Eghlid plain, Fars Province, Iran. *European Journal of Geography* 7, 48–60.
- Moreira, M. A., 2005. *Fundamentos do Sensoriamento Remoto e Metodologia de Aplicação*. 3ª edição. Viçosa. Editora UFV.
- Naue, C.R., Marques, M.W., Lima, N.B., Galvêncio, J.D., 2011. Sensoriamento remoto como ferramenta aos estudos de doenças de plantas agrícolas: uma revisão (Remote Sensing as a Toll for the Study of Plant Diseases on Agriculture: a Revision). *Revista Brasileira de Geografia Física* 3, 01-190.
- Novo, E. M. L. M., 2010. *Sensoriamento Remoto: princípios e aplicações*. 2ª edição. São Paulo: Edgard Blücher, 01-308.
- Novo, E. M. L. M., Ponzoni, F. J., 2001. *Introdução ao sensoriamento remoto*. Instituto Nacional.
- Nutter, F.W., Tylka, G.L., Guan, J., Moreira, A.J.D., Marett, C.C., Rosburg, T.R., Basart, J.P., Chong, C.S., 2002. Use of remote sensing to detect soybean cyst nematode-induced plant stress. *Journal of Nematology* 34, 222–231.
- Oliveira, T.H., Galvencio, J.D., 2009. *Caracterização Ambiental Da Bacia Hidrográfica*

- Do Rio Moxotó – Pe Usando Sensoriamento Remoto Termal. *Revista Brasileira de Geografia Física* 1, 30.
- Pinter, P.J., Ritchie, J.C., Hatfield, J.L., Hart, G.F. 2003. The Agricultural Research Service's remote sensing program: An example of interagency collaboration. *Photogrammetric Engineering and Remote Sensing* 69, 615–618.
- Poças, I., Cunha, M., Pereira, L.S., 2012. Dynamics of mountain semi-natural grassland meadows inferred from SPOT-VEGETATION and field spectroradiometer data. *International Journal of Remote Sensing* 4334–4355.
- Ponzoni, F. J., Disperati, A. A., 2001. Comportamento espectral da vegetação. *Sensoriamento remoto: reflectância de alvos naturais*, 1, 157-199.
- Ponzoni, F.J., Shimabukuro, Y., Kuplich, T. M., 2012. *Sensoriamento Remoto da Vegetação*, 2^a ed., 160.
- Rey-Caramés, C., Diago, M.P., Pilar Martín, M., Lobo, A., Tardaguila, J., 2015. Using RPAS multi-spectral imagery to characterise vigour, leaf development, yield components and berry composition variability within a vineyard. *Remote Sensing* 7, 14458–14481.
- Richards, J. A.; Jia, X., 2006. *Remote Sensing Digital Image Analysis: An Introduction*. ed.4a.01-454.
- Riedell, W. E., Blackmer, T. M. 1999. Leaf reflectance spectra of cereal aphid damage wheat. *Crop Science*, 1835–1840.
- Rodrigues, M.T., Rodrigues, B.T., Campos, S., 2015. Desempenho Da Classificação Supervisionada Em Diferentes Sistemas De Informação Geográfica. *Periódico Eletrônico Fórum Ambiental da Alta Paulista* 11, 25–37.
- Rodriguez-Galiano, V., Chica-Olmo, M., 2012. Land cover change analysis of a Mediterranean area in Spain using different sources of data: Multi-seasonal Landsat images, land surface temperature, digital terrain models and texture. *Applied Geography* 35, 208–218.
- Russel, S., Norvig, P., 2013. *Artificial intelligence: a modern approach*. Pearson Education Limited.
- Salama, M.S., Radwan, M., van der Velde, R., 2012. A hydro-optical model for deriving water quality variables from satellite images (HydroSat): A case study of the Nile River demonstrating the future Sentinel-2 capabilities. *Physics and Chemistry of the Earth* 50–52, 224–232.
- Schowengerdt, R.A., 2006. *Remote sensing: models and methods for image processing*. Elsevier.
- Shimabukuro, Y.E., Ponzoni, F.J., 2012. Orbital sensors data applied to vegetation studies. *Revista Brasileira de Cartografia* 4, 873–886.

- Shiratsuchi, L. S.; Christoffoleti, P. J.; Fontes, J. R. A., 2003. Aplicação localizada de herbicidas. *Embrapa Cerrados-Documentos (INFOTECA-E)*, 01–18.
- Sibanda, M., Mutanga, O., Rouget, M., 2015. Examining the potential of Sentinel-2 MSI spectral resolution in quantifying above ground biomass across different fertilizer treatments. *ISPRS Journal of Photogrammetry and Remote Sensing* 110, 55–65.
- Sims, D.A., Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment* 81, 337–354.
- Song, J., Li, S., Wei, W., Xu, Y., Yao, Q., 2017. Assessment of parasitic fungi for reducing soybean cyst nematode with suppressive soil in soybean fields of northeast China. *Acta Agriculturae Scandinavica, Section B—Soil & Plant Science*, 67, 730-736.
- Taghizadeh-Mehrjardi, R., Nabiollahi, K., Minasny, B., Triantafyllidis, J., 2015. Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region, Iran. *Geoderma* 253–254, 67–77.
- Thenkabail, P. S., Lyon, J. G. (Eds.), 2016. *Hyperspectral remote sensing of vegetation*. CRC press.
- Tran, D.X., Pla, F., Latorre-Carmona, P., Myint, S.W., Caetano, M., Kieu, H. V., 2017. Characterizing the relationship between land use land cover change and land surface temperature. *ISPRS Journal of Photogrammetry and Remote Sensing* 124, 119–132.
- Van der Werff, H., van der Meer, F., 2015. Sentinel-2 for mapping iron absorption feature parameters. *Remote Sensing* 7, 12635–12653.
- Vieira, S.R., Gonzalez, A.P., 2003. Analysis of the spatial variability of crop yield and soil properties in small agricultural plots. *Bragantia* 62, 127–138.
- Witten, I.H., Frank, E., 2002. *Data Mining: Practical Machine Learning Tools and Techniques*. *Acm Sigmod Record*, 31, 76-77.
- Wu, C., Niu, Z., Tang, Q., Huang, W., 2008. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and Forest Meteorology* 148, 1230–1241.
- Zhang, M., Qin, Z., Liu, X., Ustin, S.L., 2003. Detection of stress in tomatoes induced by late blight disease in California, USA, using hyperspectral remote sensing. *International Journal of Applied Earth Observation and Geoinformation* 4, 295–310.
- Zhang, H., Song, Q., Griffin, J.D., Song, B.H., 2017. Genetic architecture of wild soybean (*Glycine soja*) response to soybean cyst nematode (*Heterodera glycines*). *Molecular Genetics and Genomics* 292, 1257–1265.

JUSTIFICATIVA

O reconhecimento inicial observado por produtores de soja quanto ao ataque de nematoides na cultura da soja é tipicamente realizado pela observação visual do sistema radicular das plantas. O custo de coleta de dados para determinar populações de nematoides no campo inviabiliza amostragens em larga escala, e pode resultar em manejo inadequado em algumas propriedades rurais. O sensoriamento remoto é uma alternativa promissora para o mapeamento e identificação de nematoides, direcionando através de imagens os locais exatos que estão infectados pelos nematoides. A integração de técnicas laboratoriais com a classificação de imagens pode reduzir os custos e aumentar a eficiência do manejo de nematoides. Outra vantagem é que as técnicas de aprendizado de máquina aplicado as imagens orbitais podem utilizar dados de diferentes formatos e com funções de densidade de probabilidade desconhecidas (Ali et al., 2015) . Poucos estudos têm avaliado a generalização de algoritmos de aprendizado de máquina no decorrer do ciclo produtivo, em que imagens registradas em uma época são utilizadas para classificar cenas separadas no tempo. A abordagem multitemporal pode auxiliar no mapeamento de alterações espectrais da soja, além de determinar a dispersão e distribuição espacial de nematoides. Portanto, determinar a robustez do algoritmo baseado em aprendizagem de máquina aplicado a imagens de satélites podem auxiliar na classificação e identificação de áreas infestadas por nematoides atacando a cultura da soja.

OBJETIVO GERAL

O objetivo geral deste estudo foi avaliar o potencial do algoritmo *Random Forest* derivados de aprendizado de máquina para classificação de imagens orbitais utilizadas no monitoramento sazonal de *P. brachyurus* por meio da classificação de imagens do sensor *Multispectral Instrument* (MSI) a bordo dos satélites Sentinel-2A e Sentinel-2B.

Objetivos Específicos

- Avaliar o potencial do classificador Random Forest para identificar características espectrais dos agrupamentos de plantas infestadas (reboleiras) de *P. brachyurus* ao longo do ciclo da cultura de soja;
- Caracterizar a evolução e o comportamento espectro-temporal de infestações de *P. brachyurus* em áreas de soja;
- Identificar índices de vegetação que possam otimizar a detecção espectral de *P. brachyurus* usando a reflectância das plantas de soja nas faixas do visível e infravermelho-próximo;
- Determinar a viabilidade de uso dos índices de vegetação do visível para monitoramento de reboleiras de nematoides.

CAPÍTULO ÚNICO

(Normas de acordo com a revista “ISPRS Journal of Photogrammetry and Remote Sensing”)

Algoritmo de floresta aleatória para classificar plantas de soja infestadas por *Pratylenchus brachyurus* a partir de imagens de satélite

Random forest algorithm to classify soybean plants infested by *Pratylenchus brachyurus* using satellite images.

Resumo: O Brasil é o maior produtor mundial de soja. A infecção por fitonematoites na cultura da soja causa grandes perdas de produtividade. Considerando que as plantas atacadas por nematoides têm seu comportamento espectral alterado em relação às plantas saudáveis, este estudo objetivou determinar o potencial do algoritmo classificador Random forest para a identificação de agrupamentos de plantas infectadas por *P. brachyurus* em soja a partir de imagens coletadas pelo satélite Sentinel-2. Foi realizado um voo com um RPA multirrotor na área com histórico de infestação de nematoides. Foram identificadas 5 zonas (reboleiras) com plantas que apresentaram características visíveis com sintomas de *P. brachyurus*, 9 pontos de coleta de solo e raiz foram definidos em cada reboleira, totalizando 45 pontos com presença de nematoides. As coletas foram feitas em um padrão específico de distâncias de 10, 20, 40 e 80 m do centro das reboleiras. As imagens orbitais foram adquiridas gratuitamente pela plataforma Agência Espacial Europeia - ESA e pela plataforma de armazenamento e processamento em nuvem do Google Earth Engine no qual foram calculados 13 índices de vegetação e definido 50 pontos de pseudoausência de nematoides escolhidos aleatoriamente sobre a área de interesse para as datas durante o ciclo da cultura da soja na safra de 2018/2019 disponíveis. Os dados foram subdivididos em 4 épocas (Época 1 são imagens registradas no início de semeadura até o desenvolvimento do terceiro trifólio totalmente expandido aproximadamente (V3), Época 2 é composta por imagens até o final do estágio vegetativo aproximadamente, Época 3 são imagens registradas do início de semeadura até aproximadamente o estágio reprodutivo e a época 4 é composta por imagens referentes a todo o ciclo da cultura de soja), a exatidão da classificação foi avaliada por exatidão global e índice Kappa. Os resultados a partir do modelo gerado para época 1 apresentaram exatidão de 0,78 e índice Kappa de 0,48, para o modelo da época 2 a exatidão foi de 0,83 e índice Kappa de 0,59, para época 3 precisões de 0,82 e índice Kappa de 0,56 e por último a época 4 constituída por todas as variáveis disponíveis apresentando como melhor valor de exatidão de 0,89 e índice Kappa de 0,75. Quanto aos mapas temáticos gerados após a classificação, percebe-se que o modelo da época 3 e época 4 apresentaram produtos semelhantes, no que se refere à definição e tamanho das áreas contaminadas, essa semelhança existe devido à reflectância na banda do vermelho e do infravermelho das variáveis inseridas no quarto modelo não terem sido significativas em relação às variáveis do terceiro modelo. Assim, a técnica de classificação pelo *Random Forest* mostrou eficiência na detecção de agrupamentos de plantas com nematoides em soja.

Palavras-chave: sensoriamento remoto, sentinel 2, machine learning, fitonematoide, *glycine max L.*

Abstract: Brazil is the world's largest soybean producer. Infection with phytonemus in soybean culture causes great loss of productivity. Considering that the plants attacked by nematodes have their spectral behavior altered in relation to healthy plants, this study aimed to determine the potential of the Random forest classifier algorithm for identification of groups of plants infected by *P. brachyurus* in soybean from images collected by the Sentinel satellite -2. A flight was carried out with a multicopter RPA in the area with a history of nematode infestation. Five zones were identified with plants that showed visible characteristics with symptoms of *P. brachyurus*, 9 points of soil and root collection were defined in each zone, totaling 45 points with the presence of nematodes. The collections were made in a specific pattern at distances of 10, 20, 40 and 80 m from the center of the reeds. The orbital images were acquired free of charge by the European Space Agency - ESA platform and by the Google Earth Engine cloud storage and processing platform, in which 13 vegetation indexes were calculated and 50 randomly chosen pseudo-absence points of nematodes were defined on the area for the dates during the soybean crop cycle in the 2018/2019 harvest available. The data were subdivided into 4 seasons (Season 1 are images recorded at the beginning of sowing until the development of the third fully expanded trefoil approximately (V3), Season 2 is composed of images until the end of the vegetative stage approximately, Season 3 are recorded images of beginning of sowing until approximately the reproductive stage and season 4 is composed of images referring to the entire soybean crop cycle), the classification precision was evaluated by global accuracy and Kappa index. The results from model generated for season 1 showed precision of 0.78 and a Kappa index of 0.48, for model of season 2 the precision was 0.83 and a Kappa index of 0.59, for season 3 precision of 0.82 and Kappa index of 0.56 and lastly season 4, consisting of all available variables, with the best precision value of 0.89 and Kappa index of 0.75. As for the thematic maps generated after the classification the model from season 3 and season 4 presented similar products, about to the definition and size of the contaminated areas. This similarity exists due to the reflectance in the red and infrared band of the variables included in the fourth model that were not significant in relation to the variables in the third model. Thus, the classification technique by Random Forest showed efficiency in detecting clusters of plants with soybean nematodes.

Keywords: remote sensing, sentinel 2, machine learning, *Glycine max L.*

1.Introdução

O Brasil é maior produtor mundial de soja (*Glycine max* L), com produção total na safra 2019/20 de aproximadamente 122,2 milhões toneladas, aumento de 5,1% em relação à safra 2018/19. O estado de Goiás se destaca por ser um dos maiores produtores de soja no país, com produção aproximadamente 12 milhões de toneladas. No sudoeste do estado, o município de Rio Verde tem se destacado na produção de grãos com 1,1 milhões de toneladas (CONAB, 2020). O aumento no volume de soja produzido pelo país esteve associado principalmente as condições climáticas favoráveis para produção agrícola e a implementação de novas tecnologias (FIESP; CONAB, 2020). No entanto, ainda existem fatores que comprometem a produtividade, tais como: presença de plantas daninhas, ataque de insetos-praga e doenças, deficiência nutricional, estresse hídrico e, principalmente, os nematoides. Além de perdas em produtividade, os nematoides são problemas fitossanitários que podem levar a inviabilização de área agrícolas para plantio de outras culturas.

As perdas de produtividade causadas pela infestação de nematoides em 14 países e 40 culturas foram estimados em 13,5%, resultando em prejuízos anuais de 358 bilhões de dólares (Franchini, 2014). No Brasil, as perdas na soja são estimadas em 53,3 bilhões de dólares ao ano (SBN, 2017). As espécies de maior importância econômica para soja são os nematoides da espécie *Heterodera glycines* (nematóide de cisto), *Pratylenchus brachyurus* (nematóides das lesões radiculares) e *Meloidogyne* spp. (nematóides de galhas), vale ressaltar que o nematóide-das-lesões-radiculares (*Pratylenchus brachyurus*) tem causado danos econômicos elevados à cultura da soja no Brasil, principalmente na região Centro-Oeste, com perdas de produtividade de até 50% (Franchini et al., 2014).

O método tradicional para identificação das espécies de nematoides em campo utiliza observação visual do sistema radicular, que pode ou não apresentar sintomas típicos da presença de tais organismos, fazendo-se necessário a amostragem de raízes das plantas e solo para melhor exatidão e distribuição espacial. Os principais métodos de manejo podem reduzir a população do nematóide, tais como controle biológico, rotação

de culturas, defensivos químicos, cultivares resistentes e plantas antagonistas (Xie et al., 2016; Jardim et al., 2018).

As metodologias de manejo e controle nem sempre conseguem controlar o fator de reprodução da população de nematoides, pois esses nematoides estão localizados no solo e agem na degradação do sistema radicular da planta. Além disso, nematoides possuem capacidade de sobreviver dentro de raízes e em diferentes profundidades no solo por anos (Forge et al., 1998; Pudasaini et al., 2006). O custo de amostragem de nematoides é elevado e, por isso, torna-se necessário utilizar métodos inovadores com menor custo para caracterizar a infestação desses fitoparasita (Machado, 2015).

O desenvolvimento de relações entre as características da vegetação e dados coletados remotamente tem sido meta de muitos estudos, principalmente aplicado ao setor agrícola para identificação e detecção da presença de patógenos, através de uma série de avaliações que demonstram os comportamentos que as radiações eletromagnéticas (REM) assumem ao interagir com a cobertura vegetal (Abreu & Coutinho, 2014; Martins & Galo, 2015). Entre os métodos utilizando técnicas de sensoriamento remoto (SR), destaca-se análises da resposta espectral no dossel da planta e efeitos morfofisiológicos causados pelo estressor. Uma das medidas mais utilizadas para demonstrar essas variações são os denominados índices de vegetação que são modelos matemáticos com proposição de combinações específicas de bandas espectrais. Através desses modelos é possível explorar propriedades espectrais da vegetação relacionadas a parâmetros biofísicos, bioquímico e condições fisiológicas da cobertura vegetal (Ponzoni et al., 2007).

Foram realizados alguns estudos para monitorar, analisar, diagnosticar e mapear áreas com a ocorrência de infestações de nematoides (Martins et al., 2017), presença de fungos (*Sclerotinia sclerotiorum*) (Cao et al., 2018), ataque de pulgão da soja (Alves et al., 2015 ; Alves et al., 2019), detecção de patógenos (Joalland et al., 2018), avaliar o estado nutricional de plantas (Zheng et al., 2018), aplicações localizadas (Shiratsuchi et al., 2003), e auxiliar no desenvolvimento e implementação da agricultura de exatidão (Chemura et al., 2017; Yang, 2018; Campos et al., 2019).

Nas últimas décadas, o desenvolvimento de novas tecnologias digitais derivadas do SR vem influenciando a tomada de decisões na agricultura. Ferramentas digitais, como sistemas de informação geográfica (SIG), têm desempenhado papel importante no desenvolvimento de novos métodos de processamento e visualização de informações espaciais, bem como no monitoramento, avaliações, análises e mapeamento das lavouras

sendo possível estimar a produtividade, avaliar deficiências nutricionais, detectar insetos-praga, doenças, plantas daninhas, falha de plantio e necessidade hídrica das plantas. O sensoriamento remoto utiliza da pressuposição de que culturas fisiologicamente estressadas possuem comportamento espectral diferente das plantas saudáveis (Riedell & Blackmer, 1999; Zhang et al., 2003; Marin et al., 2019).

As variações da resposta espectral de culturas agrícolas podem ser identificadas aliadas a técnicas de SR aplicadas em campo através de sensores RGB, multiespectral, termal e hiperespectral. Esses sensores podem ser acoplados a um Sistema de Aeronaves pilotadas remotamente (RPAs - *Remotely piloted aircraft systems*) em baixa altitude, ou inseridos a satélites artificiais orbitais, ou sensores terrestres como os espectrorradiômetros.

Em virtude da alta capacidade de capturar informações espectrais do dossel das plantas a uma resolução espacial centimétrica, disponibilidade de voos em curto intervalo de tempo e por auxiliarem em melhor tomada de decisão, o uso de tecnologias RPA's para agricultura já estão sendo utilizadas por vários produtores das propriedades rurais no sudoeste Goiano. Em maior escala é importante ressaltar que para o monitoramento de grandes áreas o uso de imagens orbitais se torna oportuna, pelo desenvolvimento de tecnologias, métodos e técnicas que permitiram o avanço na obtenção de dados da superfície terrestre e, com isso, redução do tempo gasto com trabalhos de campo em locais de difíceis acessos. A partir dos dados oriundos de satélites surgiram diversas metodologias para elaboração de mapas, principalmente, no que tange a classificação de imagens de satélite. O processamento de imagens digitais pode facilitar a identificação e a extração das informações contidas nas imagens, para posterior interpretação. Assim, o fundamento deste processo se refere a um conjunto de medições de radiância obtidas em diferentes faixas do espectro eletromagnético (Jensen, 2011; Vale et al., 2018).

O interesse em usar dados obtidos pelo SR na agricultura para avaliações de doenças de plantas aumentou consideravelmente nos últimos anos. O satélite Sentinel-2A foi lançado em 2015 e 2B em 2017, ambos com o objetivo principal de avaliar a sanidade de vegetação em aplicações agrícolas e de ecossistemas que necessitem maior resolução espacial. Os satélites operam com revista de imageamento da superfície terrestre a cada 10 dias, que combinados proporcionam um ciclo de revisita de 5 dias. Possuem largura de faixa de imageamento de 290 km e bandas espectrais exclusivas projetadas para avaliação da condição da vegetação (ESA, 2019). O Sentinel-2 possui um instrumento de captura multiespectral, sendo que a câmera é a mais avançada do seu tipo, por incluir três

bandas no “red-edge” que geram informações de suma importância sobre o estado da vegetação. Portanto, a nova geração de sensores espaciais, como o MSI, pode mapear efetivamente algumas características morfológica e fisiologicamente relevantes da vegetação com base na resposta espectral do dossel de plantas (Taquia, 2015).

Os dados gerados através das técnicas de SR formam grande massa de dados digitais que necessitam de técnicas específicas para processamento, análise e interpretação. O aprendizado de máquina (*machine learning*) é uma dessas técnicas capazes de processar os dados e transformá-los em informações úteis e fáceis de interpretação, está diretamente relacionado com a classificação de imagens baseadas em algoritmos (Maxwell et al., 2018; Liakos et al., 2018). Trata-se de uma alternativa promissora para o mapeamento de culturas principalmente em relação a nematoides, devido suas características fisiológicas serem parecidas com algumas deficiências nutricionais (deficiência de nitrogênio, deficiência de potássio, deficiência de magnésio, entre outros), causando confusão espectral entre as plantas em uma única cena. Esses algoritmos possuem a capacidade de realizar mapeamentos complexos, explorando as informações contidas em um conjunto de amostras de referência (Ali et al., 2015).

Um dos métodos robustos comumente utilizados na agricultura é o Random Forest (RF) que se trata de um classificador supervisionado de aprendizado em máquina não paramétrico baseado no conceito de árvores de decisão para classificação. Esse método de aprendizado utiliza um conjunto de árvores preditoras contínuas que crescem a partir de outras árvores. Cada árvore é determinada de maneira independente utilizando uma amostra de *bootstrap* do conjunto de dados (Liaw & Wiener, 2002). Para o emprego dessa técnica, dois parâmetros devem ser definidos: o número de variáveis de entrada escolhidas aleatoriamente em cada divisão (*mtry*) e o número de árvores (*ntree*) (Breiman, 2001).

Diante disso, o objetivo deste estudo foi analisar o desempenho de bandas espectrais e índices de vegetação a partir de imagens oriundas do satélite Sentinel-2 para detectar áreas com presença de nematoides, com o emprego do algoritmo classificador paramétrico Random Forest. Este estudo tem o potencial de otimizar o uso de satélites na detecção de problemas na soja a nível espacial e temporal de infestação de nematoides e a forma como estão espalhados pelo campo.

2. Material e Métodos

2.1 Área de estudo

O estudo foi realizado em área de cultivo de soja situada no município de Rio Verde – uma região com grande quantidade de lavouras. Rio Verde está localizado na região do sudoeste do estado de Goiás, com altitude 739 m, latitude 17°47'52" Sul e longitude 51°55'40" Oeste (IBGE, 2020; Fig. 3). O clima do município é classificado como tropical, predominância de solo classificado em Latossolo Vermelho Distrófico (INPE, 2008), declividade que vai de plano (0 a 3%) a suave ondulado (3 a 8%), temperatura média de 22,3°C e precipitação anual aproximadamente de 1.600 milímetros (SOMA BRASIL, 2019). Análises preliminares determinaram ocorrência de nematoides referente as espécies de *P. brachyurus*.

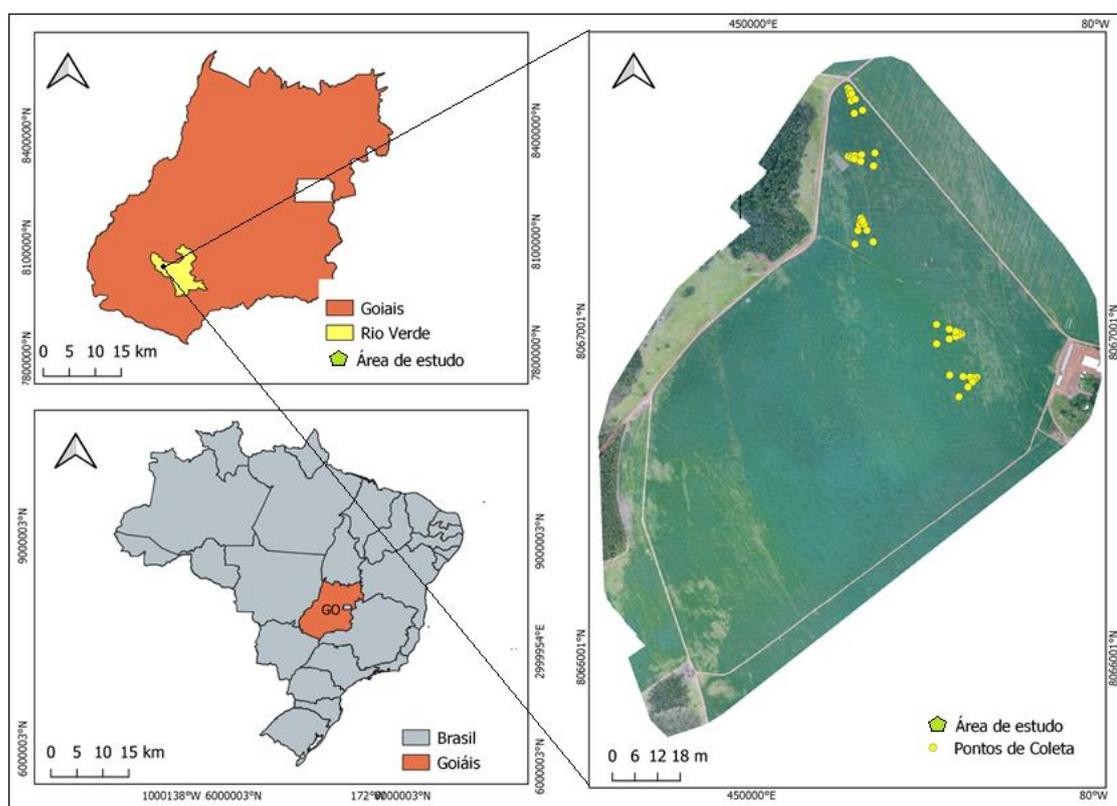


Figura 3. Localização geográfica da área de estudo no município de Rio Verde – GO

2.3 Identificação de zonas de amostragem de nematoides

O levantamento dos dados *in situ* de referência sobre a ocorrência de infestação de nematoides foi realizado no início do estágio de florescimento (R1) durante a safra 2018/2019 (Fig. 4). Um voo exploratório inicial foi realizado para identificar potenciais grupos de plantas que mostravam diferenças espectrais (coloração) utilizando um RPA

(Phantom 4 Advanced, DJI, Shenzhen, China) equipado com sensor RGB (20 MP CMOS, DJI, Shenzhen, China) a 120 m de altura e sobreposições laterais e frontais de 80% (pixel de 3,6 cm). As imagens foram ortorretificadas no software Pix4D 3.2.23 (Pix4D SA, Lausanne, Suíça).

Cinco zonas de amostragem com sintomas visíveis no sistema radicular das plantas causados por nematoides foram selecionadas durante uma inspeção. Dois transectos representando as direções aparentes da distribuição local do nematoide foram criados para cada zona usando a imagem ortorretificada no software SIG (QGIS Development Team, Boston, EUA). Os transectos foram estabelecidos encontrando as coordenadas geográficas do centro da zona danificada e posicionando radialmente pontos de amostragem adicionais a 10, 20, 40 e 80 m do centro das reboleiras (ou seja, um total de nove pontos de amostragem por zona ou 45 pontos de amostragem por campo).

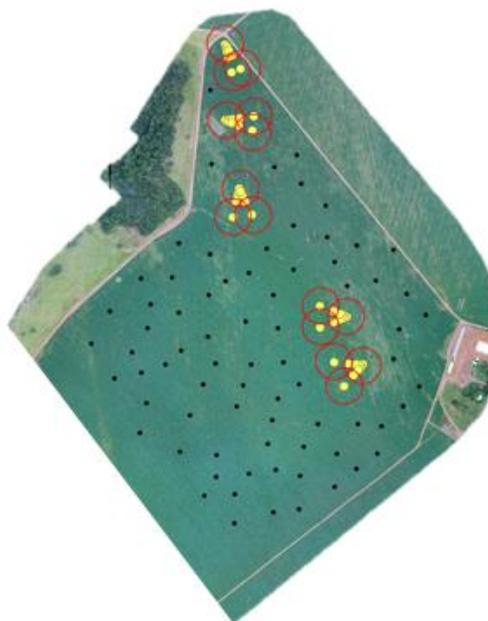


Figura 4. Representação da geometria dos pontos de coleta em campo em cada reboleira de nematoide.

2.4 Identificação de nematoides no solo e raízes

As espécies de nematoides foram identificadas por análises nematológicas realizadas pelo laboratório de fitopatologia com apoio financeiro do laboratório de VANT's na agricultura (LDVA), Polo de Inovação IF Goiano, Rio Verde, GO, na safra de 2018/2019. As amostras de solo e raiz foram obtidas em todos os 45 pontos de amostragem no estágio de florescimento das plantas de soja (R1). Três subamostras (500 g cada) na camada de 0-20 cm foram coletadas com uma pequena enxada e homogeneizadas em um recipiente. Para cada ponto de amostragem, uma amostra de solo

(500 g) foi obtida das subamostras homogêneas, armazenada em sacos plásticos com 50 g de amostras de raízes guardadas a 4°C até a identificação. A extração de *P. brachyurus* no solo e raízes foi realizada utilizando os métodos descritos por Jenkins (1964); Coolen & D'Herde (1972). Os nematoides foram identificados por espécie em comparação a uma coleção de referência e chaves taxonômicas (Costa-Manso et al., 1994).

2.5 Aquisição de imagens orbitais

As cenas adequadas para o estudo foram selecionadas a partir dos seguintes critérios: imagens correspondentes ao ciclo da cultura, ausência de nuvens e principalmente pela proximidade com a data de análises das variáveis agrônômicas (população de nematoides no solo e raiz e altura de planta) correspondente a 21/12/2018. A base de dados orbitais foi constituída com imagens referente a 3/10/2018, 18/10/2018, 7/11/2018, 12/12/2018, 17/12/2018, 11/01/2019, 16/01/2019, 31/01/2019.

Os sensores abordo do Sentinel 2A e 2B fornecem imagens coletadas com resolução radiométrica de 12 bits, possui quatro bandas com 10 m de resolução espacial, seis bandas com 20 m de resolução espacial e três bandas com 60 m de resolução espacial (Tab. 2). A aquisição das imagens é disponibilizada gratuitamente pela Agência Espacial Europeia e pela plataforma de armazenamento e processamento em nuvem do Google Earth Engine.

Tabela 2. Características das bandas multiespectrais do sensor MSI/Sentinel-2A e 2B.

Resolução espacial (m)	Número da banda	Nome da banda	Comprimento de onda central (nm)
10	B02	Blue (Azul)	490
	B03	Green (Verde)	560
	B04	Red (Vermelho)	665
	B08	NIR (Infravermelho Próximo)	842
20	B05	Red Edge 1	705
	B06	Red Edge 2	740
	B07	Red Edge 3	783
	B08 A	Red Edge 4	865
	B11	SWIR1	1610
	B12	SWIR2	2190
60	B01	Aerossol	443
	B09	Water Vapor	940
	B10	Cirrus	1375

Fonte: Adaptado de Chemura et al., 2018.

2.6 Processamento de imagens multiespectrais e índices de vegetação

Todas as imagens foram adquiridas e manipuladas em linguagem de programação Java Script para o cálculo dos índices de vegetação através da plataforma do Google Earth Engine, na forma de produto nível 1-C, ou seja, imagens geometricamente corrigidas (ortorretificada), com valores de reflectância no topo da atmosfera (ToA). Treze índices de vegetação foram calculados para cada *pixel* das imagens dentro dos pontos amostrais (Tab. 3).

Tabela 3. Índices de vegetação calculados a partir da reflectância das imagens multiespectrais do sensor Sentinel-2 MSI conforme base de dados *Index Data base*.

Índices de vegetação	Equação	Fonte
Normalized Difference Vegetation Index (NDVI)	$NDVI = (B08 - B04) / (B08 + B04)$	Rouse et al., (1973)
Normalized Difference Vegetation Index (NDVI ₆₉₀₋₇₁₀)	$NDVI_{69} = B08 - (690:710) / B08 + (690:710)$	Datt, B. (1999)
Green Leaf Index (GLI)	$GLI = (B03 - B04) + (B03 - B02) / (2 * B03) + B04 + B02$	Louhaichi et al., (2001)
Atmospheric resistant vegetation index (ARVI)	$ARVI = (B08 - (2 * B04) + B02) / (B08 + (2 * B04) + B02)$	Kaufman & Tanré (1992)
Leaf Chlorophyll Index (LCI)	$LCI = (850 - 710) / (850 + 680)$	Datt, B. (1999)
Atmospherically Resistant Vegetation Index 2 (ARVI2)	$ARVI2 = -0.18 + 1.17 * ((B08 - B04) / (B08 + B04))$	Kaufman & Tanré (1992)
Green Atmospherically Resistant Index (CHLgreen)	$CHL_{green} = ((760:800) / (540:560))^{(-1)}$	Gitelson et al., (2006)
Chlorophyll Red-Edge Index (CHLRedEdge)	$CHL_{RedEdge} = ((760:800) / (690:720))^{(-1)}$	Gitelson et al., (2006)
Chlorophyll Absorption Ratio Index (CARI)	$CARI = \left(\frac{700nm}{670nm} \right) \left(\frac{\sqrt{(a - 670 + 670nm + b)^2}}{(a^2 + 1)^{0.5}} \right)$	Kim et al., (1994)
Chlorophyll Absorption Ratio Index 2 (CARI2)	$CARI2 = ((a - 670) + (b + 670) / (a^2 + 1)^{0.5}) * (700/670)$	Kim et al., (1994)
Canopy Chlorophyll Content Index (CCCI)	$CCCI = ((B08 - B05 / B08 + B05) / (B08 - B04 / B08 + B04))$	El-Shikha et al., (2008)
Chlorophyll vegetation index (CVI)	$CVI = B08 * (B04 / B03^{(2)})$	Gornitz (1991)

Modified Chlorophyll Absorption in Reflectance Index / Modified Triangular Vegetation Index 2 (MCARI/MTVI2)	$\frac{\text{MCARI}}{\text{MTVI2}} \left(\frac{(700 - 670) - 0.2(700 - 550) \left(\frac{700}{670}\right)}{1.5 \frac{1.2(800\text{nm} - 550\text{nm}) - 2.5 (670\text{nm} - 550\text{nm})}{\sqrt{(2800\text{nm} + 1)^2 - 6800\text{nm} - 5\sqrt{670\text{nm}} - 0.5}}} \right)$	Kim et al., (1994)
---	---	-----------------------

Em que: IV: índice de vegetação; B02: *blue* (490 nm); B03: *Green* (560 nm), B04: *red* (665 nm); B05: *RedEdge* (705 nm); B08: *NIR*-infravermelho próximo (842 nm); a: (700nm-550nm) / 150; b: (550nm-((700nm-550nm)/150*550)).

2.7 Algoritmos de análise e classificação de dados

A classificação das imagens foi feita através do método Random Forest, executado no pacote estatístico R pela interface do *RStudio*, com base no pacote *random Forest* (Liaw & Wiener, 2002) e acessada através do pacote *caret*. O *RStudio* é um software de código aberto e livre, além de poder ser utilizado para realizar atividades computacionais com possibilidade de manipulação das configurações dos algoritmos de classificação desde que compatível com suas capacidades, realiza tarefas como comparação com base de dados, processamento de dados, classificação, regressão, agrupamentos, regras de associação e visualização, modelagem, produção de mapas e gráficos e outros (Kuhn, 2017).

Cada árvore do algoritmo Random Forest gerou um voto que indica a decisão sobre a escolha da classe de uma amostra, assim, a classe que obteve o maior número de votos foi atribuída a mesma. A classificação com RF depende apenas de dois parâmetros para gerar um modelo de predição: o número de árvores de decisão (*n tree*) e número de variáveis preditoras (*m try*). Para *n tree* foi utilizado o valor de 500 e número de variáveis preditoras (*m try*) escolhidas aleatoriamente em cada divisão da árvore foi ajustado de acordo com o melhor valor encontrado pelo índice Kappa (Tab. 4).

A exatidão pode melhorar ainda mais se cada árvore for treinada com seu próprio subconjunto de variáveis escolhidas aleatoriamente (*m try*) e usando apenas um subconjunto total dessas variáveis (Pal, 2005). Uma árvore treinada com um menor subconjunto de *m try* pode ter reduzida a sua força, mas é preferível que se use um pequeno número de *m try* e maior valor para *n tree* para a redução dos erros de generalização e a correlação entre as árvores, tornando o conjunto como um todo mais confiável (Breiman, 2001).

As variáveis preditoras eram constituídas de 13 índice de vegetação a partir de imagens em diferentes estádios da cultura de soja, devido essa variação foi determinado uma divisão dos dados em épocas (Tab. 4). A época 1 e época 2 foram compostas por

imagens no período vegetativo, para a época 3 foram adicionadas ao modelo imagens no estágio reprodutivo e para classificação da época 4 foram integradas todas as imagens disponíveis durante todo o ciclo da soja.

Tabela 4- Definição do número de variáveis predictoras (*mtry*) para cada época de classificação.

Época	Datas das cenas	<i>mtry</i>	Exatidão (%)	Índice Kappa
1	3 e 18/10/2018	2	77	0,41
		14	78	0,48
		26	75	0,43
2	3; 18/10/2018 e 7/11/2018	2	80	0,51
		20	82	0,59
		39	83	0,59
3	3; 18/10/2018; 7/11/2018; 12 e 17/12/2018	2	82	0,56
		33	80	0,55
		65	81	0,58
4	3 e 18/10/2018; 7/11/2018; 12 e 17/12/2018; 11, 16, e 31/01/2019	2	89	0,75
		53	84	0,65
		104	85	0,64

Em que: *mtry*= P/3 (parâmetro para classificação); Índice Kappa: pode variar de -1 discordância completa a +1 concordância completa.

2.8 Treinamento e previsões dos modelos

Considerando que o objetivo foi classificar e analisar imagens obtidas para visualizar as tendências de agrupamento de plantas e para avaliar a compatibilidade com as informações de campo da infestação de *P. brachyurus* foram considerados os pixels correspondentes aos 45 pontos de ocorrência de infestação de nematoides validados pela amostragem de campo e além desse, foram sorteados 90 pontos aleatórios de pseudoausência com distância de 50 m dos pontos de ocorrência de infestação. Esses pontos de pseudoausência foram necessários devido o modelo de classificação conter apenas pontos de presença de infestação de nematoides (Vanderwal et al., 2009; Lobo & Tognelli, 2011).

Para detectar a probabilidade de ocorrência com base nos dados de infestação de nematoide obtidos em campo e os pontos de pseudoausência de nematoides, foi realizado uma análise dos pixels de cada árvore de decisão gerada (*n tree*) em que se realizou a votação majoritária entre todas as árvores para prever a classe de presença (1) ou ausência (0) de nematoide. Uma vez que o método RF é baseado na contagem de votos das *n* árvores de decisões estimada, se em 500 árvores de decisões geradas, um determinado pixel for classificado em 400 dessas com a presença do nematoide (i.e., 400/500), logo a probabilidade de sua ocorrência foi de 0,8 em uma escala de 0 a 1.

Todos os 13 índices de vegetação foram testados como variáveis divididas em cada nó. Obteve-se 104 variáveis derivadas desses 13 índices de vegetação com datas distintas que foram utilizadas como dados de treinamento incluindo 135 pontos amostrais com duas classes (0: pseudoausência e 1: presença de nematoide) em cada variável para prever a ocorrência de nematoides da área. Estes dados de treinamento foram cuidadosamente amostrados para que não se perdesse as variações espectrais das plantas de soja. Após determinada todas as variáveis preditoras para a classificação, através da metodologia do algoritmo classificador Random Forest, obteve-se 500 árvores de decisão. Dietterich (2000) desenvolveu uma técnica de seleção aleatória em que as árvores são construídas usando um subconjunto aleatório das variáveis preditoras principais, para cada árvore. Dessa forma, cada modelo construído foi então usado para gerar previsão para uma nova amostra, essas n previsões são calculadas como média para fornecer a classificação final.

2.9 Importância de cada variável através do índice de Gini

As variáveis mais importantes para o modelo foram determinadas através do índice de Gini (ou *Mean Decrease in Impurity* – MDI), considerando a importância dos melhores índices de vegetação e as melhores datas foram classificadas pelo modelo baseados em árvore, uma métrica associada à importância de variáveis (Breiman, 2001). A importância da variável é dada pelo melhor valor encontrado, ou seja, resultados mais puros com menor índice de confusão nos resultados, quando se permuta aleatoriamente os n valores da variável e os dados originais, isto mantido fixo os valores das demais variáveis. Dessa forma, o índice de Gini se torna importante por sua capacidade de seleção das variáveis mais importantes que terão maior impacto no desempenho do modelo de classificação (Breiman & Cutler, 2003).

2.10 Avaliação da exatidão dos modelos classificados

Devido a probabilidade de se classificar incorretamente um pixel selecionado aleatoriamente existem alguns procedimentos que minimizam esse erro, assim, para a validação e calibração do modelo de classificação foi utilizado a validação cruzada de 10-*folds*, em que 9 desses foram usados para calibrar o modelo (gerar árvores de decisão) e 1 foi usado para a validação. Foi realizado esse procedimento 10 vezes de tal modo que todos os subconjuntos fossem usados para calibrar e validar o modelo, resultando em 10 matrizes de confusão (Congalton, 1991).

Por meio desta matriz foi possível efetuar a avaliação de exatidão global através do índice Kappa, esse é utilizado para verificar a confiabilidade dos mapas temáticos que são gerados a partir do processamento digital da imagem de sensoriamento remoto. Os resultados da estatística Kappa foram classificados pelo método de Landis & Koch, 1977. A exatidão global foi calculada dividindo a soma dos valores da diagonal principal da matriz, que representa o número de elementos classificados corretamente (Congalton & Green, 2019).

2.11 Produção de mapas temáticos para cada modelo

O Software QGIS 3.4.15 (QGIS Development Team, Boston, EUA) foi utilizado para geração dos mapas temáticos oriundos das classificações. Plantas cloróticas observadas nos pixels da imagem foram classificadas como probabilidade de ocorrência de nematoides como plantas com alto estresse na cor vermelha, plantas com início de sintomas de injúria causada pelos nematoides foram classificadas como plantas com baixo estresse na cor amarelo e nas plantas classificadas em verde são plantas denominadas sadias, ou seja, não foram observados sintomas visíveis de danos causados por nematoides.

3. Resultados

3.1 Parametrização de Random Forest

O desempenho da classificação de imagens utilizando o classificador *Random Forest* dependeu da disponibilidade de imagens com baixo percentual de nuvens durante o ciclo da soja e dos valores de *ntree* e *mtry* (Tabela 5). O número de *ntree* de 500 foi definido como padrão para a classificação. As variáveis de entrada fornecidas ao algoritmo para a classificação consistiram em 13 índices de vegetação calculados em diferentes datas, o valor de *mtry* de cada modelo determinado a partir da melhor exatidão encontrada (Tabela 5). A exatidão de classificação variou de 78% a 89%, com melhores índices Kappa quando mais imagens em que o ciclo da soja estava mais avançado foram inseridas no modelo (Tabela 5). Para época 1, os pixels que representavam plantas doentes e saudáveis foram classificados corretamente 78% das vezes, mesmo com nível de incidência e sintomas foliares visíveis baixos. A época 2 e época 3 apresentaram exatidão dos seus respectivos modelos de classificação bem próximas, acima de 80%. O modelo de classificação da época 4 utilizou todo o conjunto de dados, ou seja, o modelo foi

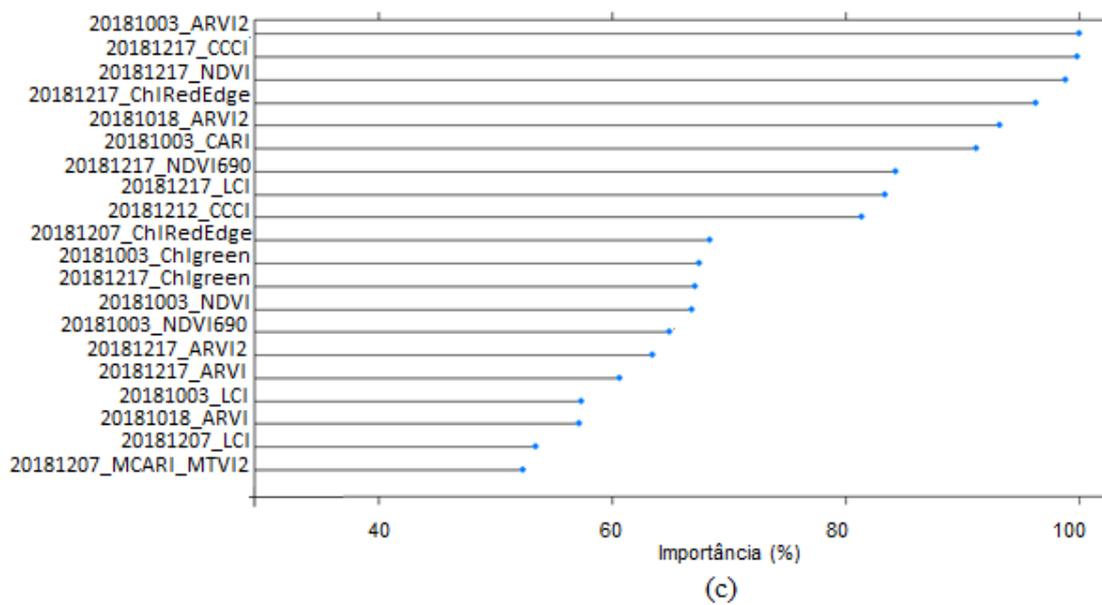
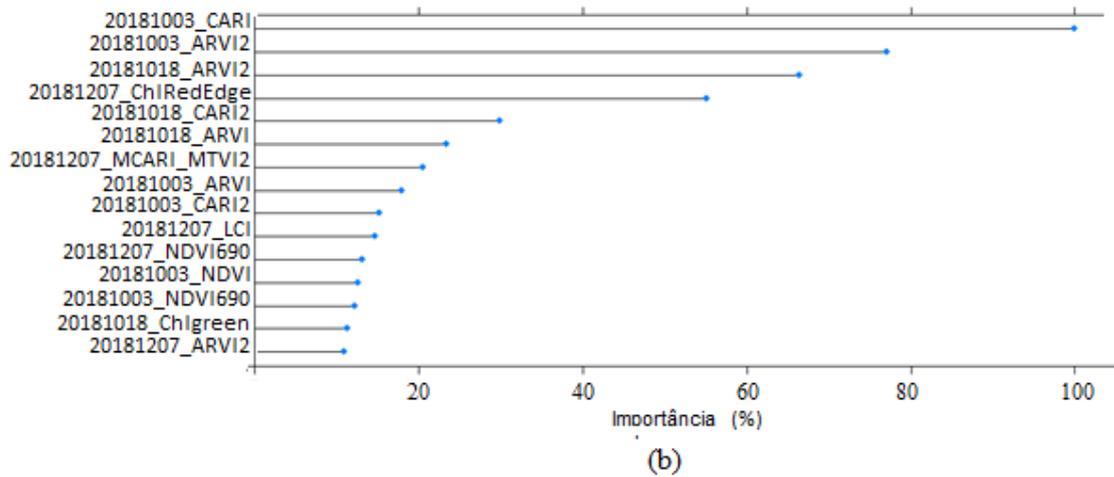
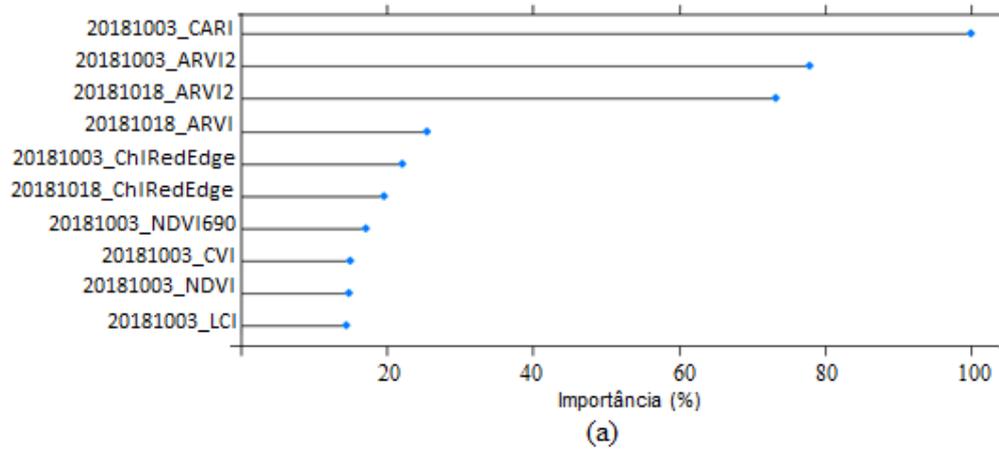
construído com dados derivados de cenas durante todo o ciclo da cultura e alcançou os melhores resultados, com exatidão de 89%.

Tabela 5 – Datas das cenas utilizadas para cada modelo, número de variáveis preditoras, parâmetro de classificação (*mtry*), exatidão e índice Kappa de modelos de classificação utilizando *Random Forest*.

Época	Datas de cenas	Variáveis preditoras (<i>n</i>)	<i>mtry</i>	Exatidão (%)	<i>Kappa</i>
1	3 e 18/10/2018	26	14	78	0,48
2	3 e 18/10/2018; 7/11/2018	39	39	83	0,59
3	3 e 18/10/2018; 7/11/2018; 12 e 17/12/2018	65	2	82	0,56
4	3 e 18/10/2018; 7/11/2018; 12 e 17/12/2018; 11, 16 e 31/01/2019	104	2	89	0,75

3.2 Importância preditiva das variáveis de entrada

A técnica de classificação utilizando o algoritmo Random Forest permitiu calcular uma métrica de importância de cada variável preditoras do modelo individualmente através do índice de Gini, dessa forma, foi possível medir a contribuição de cada uma para a exatidão da classificação realizada em cada época. Essa importância é representada na figura abaixo (Fig. 5).



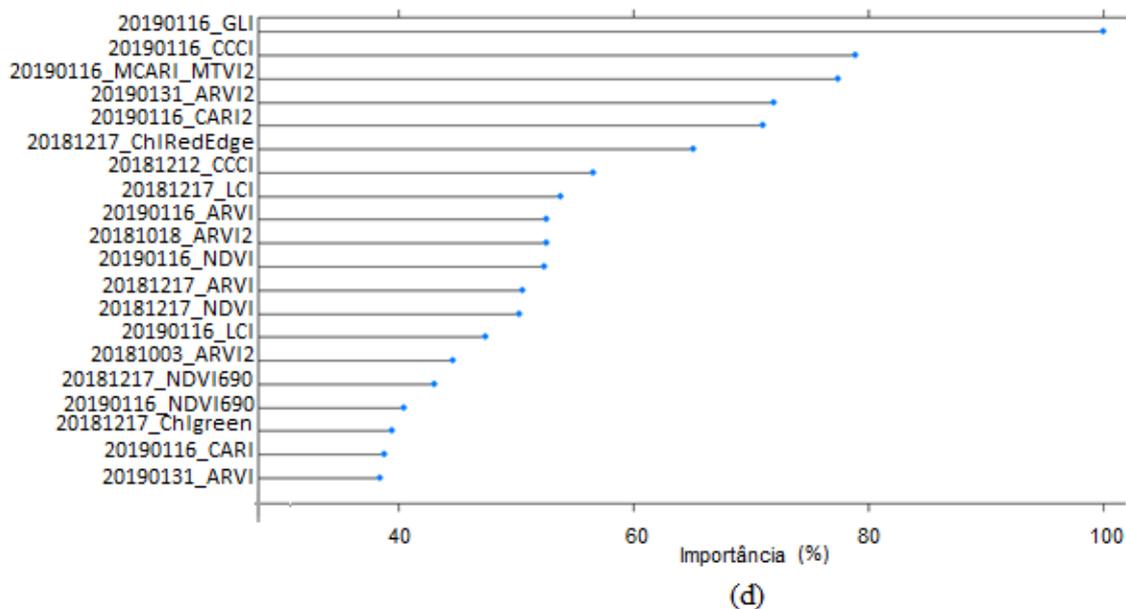


Figura 5 - Variáveis mais importantes para cada modelo classificado por Importância de Gini, considerando a importância dos melhores índices de vegetação e as melhores datas. Época 1 (a), época 2 (b), época 3 (c) e época 4 (d)

A classificação por índice de Gini mostrou para época 1 em que as plantas de soja estavam na fase inicial de desenvolvimento vegetativo com até aproximadamente 3 trifólios totalmente expandidos (V3) o índice CARI em 3/10/2018 foi o de maior relevância para classificação de reboleiras de *P. brachyurus* em soja mais do que os outros índices de vegetação calculados em outras datas (Fig. 5a). O índice CARI é derivado de comprimentos de onda na região do visível mais próximos a banda do vermelho. Um outro índice que também se mostrou importante foi o ARVI2 calculado para essa mesma data e para a cena registrada no dia 18/10/2018, porém, apresentou menor relevância quando comparado ao índice CARI (Fig. 5a). A principal diferença entre estes índices de vegetação é que CARI utiliza comprimentos de onda apenas na região do visível, portanto, pode ser replicado em imagens simples RGB, enquanto o ARVI2 é uma combinação de bandas do vermelho e do NIR.

Para a construção da modelo época 2 foi adicionado ao conjunto de variáveis os índices calculados na data 7/11/2018. Através do cálculo de importância (índice de Gini) os dados adicionados ao modelo não apresentaram maior importância/contribuição para a classificação em relação ao conjunto de dados utilizados anterior, dessa forma, para o modelo classificado na época 2, as variáveis mais importantes com resultados mais puros, ou seja, menor índice de confusão nos resultados permaneceram sendo os índices CARI e ARVI2 (Fig. 5b). Vale ressaltar que mesmo a ordem de importância não tenha sido

alterada a inclusão dessas variáveis contribuiu para que a exatidão e o valor de índice Kappa do modelo aumentasse.

A partir do subconjunto de dados da época 3, observou-se nove variáveis classificadas pela importância para o modelo de classificação em ordem crescente lugar como ARVI2 em 3/10/2018, CCCI em 17/12/2018, NDVI em 17/12/2018, ChlRedEdge em 17/12/2018, ARVI2 em 18/10/2018, CARI em 3/10/2018, NDVI_{690:710} em 17/12/2018, LCI em 17/12/2018 e, por último, CCCI em 12/12/2018 (Fig. 5c). Portanto, a cena do dia 17/12/2018 foi a que mais se destacou e teve melhor relação com vários índices de vegetação.

O modelo de classificação para a época 4 gerado pelo classificador Random Forest incluiu todo o conjunto de variáveis que foi possível obter através das imagens disponíveis pelo sensor MSI/Sentinel-2A e 2B. Nesse conjunto de dados as variáveis mais importantes foram o índice GLI, CCCI, MCARI/MTVI2, ARVI2 e CARI2 todos calculados para a mesma data de 16/01/2019 (Fig. 5d).

3.3 Avaliação de exatidão geral de cada modelo

A matriz de confusão com validação cruzada foi construída a partir de 135 pontos amostrais com pseudoausência e presença de nematoide em 4 épocas de avaliação, foi calculada para avaliar o desempenho do classificador Random Forest em detectar agrupamentos de plantas doentes. O modelo para época 1 apresentou exatidão de 77%, essa foi alcançada usando dados de cenas coletadas antes mesmo que os sintomas foliares fossem visíveis, para época 2 e época 3 com exatidão de 82% e para época 4 com exatidão de 89%.

3.4 Comparação entre exatidão e qualidade de índice Kappa de cada modelo gerado

Todo o conjunto de dados foi subdividido em 10 subconjuntos (10-folds), em que 9 foram usados para calibrar o modelo e 1 para validar, todos os subconjuntos foram usados para calibração e validação. A partir das 10 matrizes de confusão geradas, obteve-se o valor médio de exatidão e índice Kappa para cada época classificada. O modelo classificado para época 1 apresentou em média 80% de exatidão e um índice Kappa de 0,528. Para o modelo referente à época 2 em média 84% de exatidão e índice Kappa de 0,660. À época 3 uma média de 84% de exatidão e índice Kappa de 0,670, e para a última época a média de 92% de exatidão e índice Kappa maior que os anteriores (0,818).

A partir da avaliação de desempenho global através do índice Kappa, foi possível verificar a confiabilidade dos mapas temáticos (Fig.7) que foram gerados a partir da modelagem de probabilidade de ocorrência de nematoides predita pelo classificador Random Forest.

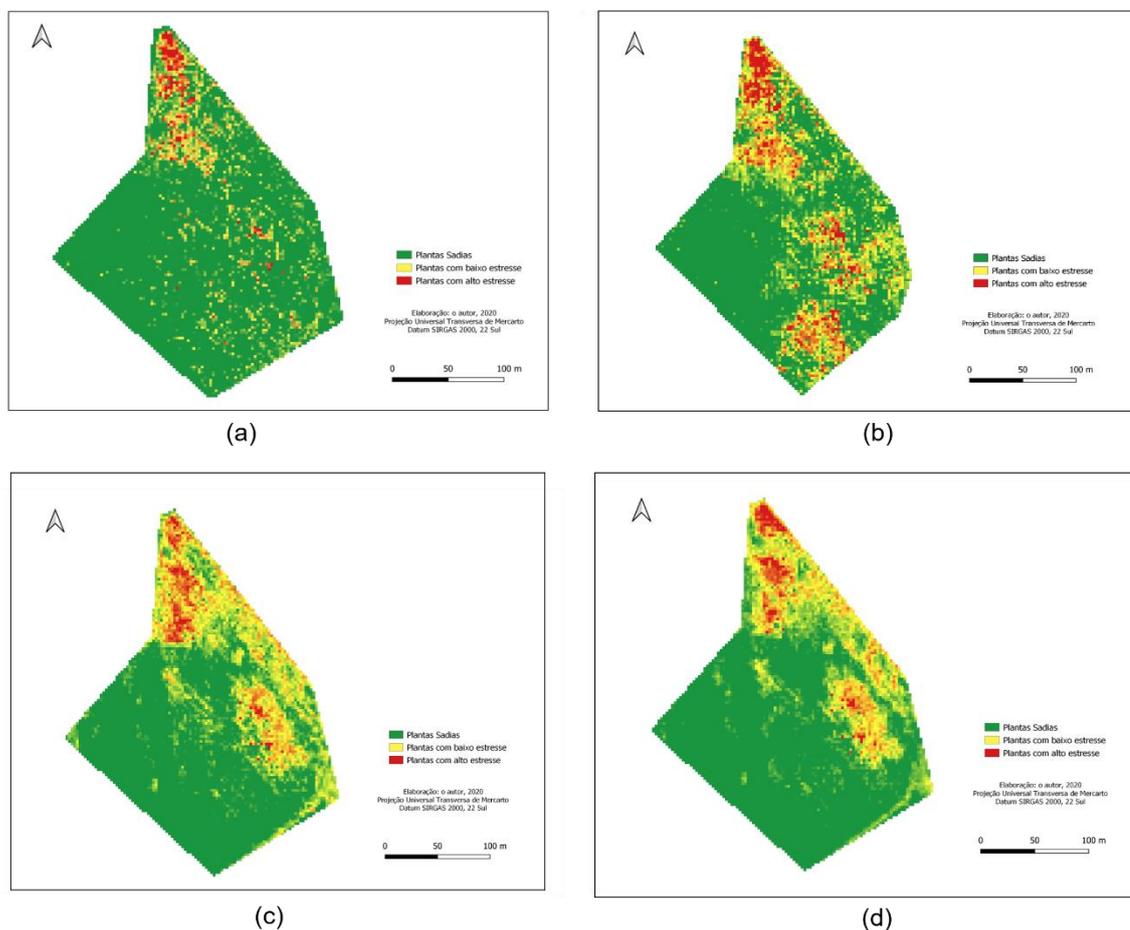


Figura 7 - Classificação da área de estudo para identificação de ocorrência de nematoides referente as espécies de *Pratylenchus brachyurus*, obtida pelo classificador Random Forest utilizando imagens orbitais disponíveis pelo sensor MSI/Sentinel-2A e 2B para época 1 (a), época 2 (b), época 3 (c) e época 4 (d). Em que: Época 1 são imagens registradas no início de semeadura até o desenvolvimento do terceiro trifólio totalmente expandido aproximadamente (V3), Época 2 é composta por imagens até o final do estágio vegetativo aproximadamente, Época 3 são imagens registradas do início de semeadura até aproximadamente o estágio reprodutivo e a época 4 é composta por imagens referentes a todo o ciclo da cultura de soja.

Foi possível identificar agrupamentos de plantas com leve início de sintomas causados por nematoide em alguns dos pontos validados em campo (Fig. 7a). Maior agrupamento de plantas infectadas com maior incidência da doença pode ser observado na imagem (b) (Fig. 7b). Observa-se também regiões classificadas com sintomas causados por nematoides em diferentes pontos de validação em campo. Deve-se considerar que este resultado pode ser decorrente da similaridade espectral de plantas atacadas por nematoides com outros estressores (Blevins et al., 1995). Em contrapartida, a “confusão” caracterizada, principalmente pela classificação de pixels de áreas que

apresentavam alguma deficiência ou estressor foram classificadas com exatidão. Um melhor ajuste do modelo foi obtido com mais variáveis preditoras adicionadas aos modelos (Figs. 7c e 7d).

4. Discussão

Dados derivados de sensoriamento remoto tem potencial para detectar mudanças espectrais causadas por nematoides no dossel da cultura de soja. Os resultados mostraram que o algoritmo classificador Random Forest conseguiu classificar agrupamento de plantas infectadas por *P. brachyurus* antes de sintomas foliares visualmente detectados na lavoura. A fim de verificar o potencial do classificador RF para predição de agrupamentos de plantas infectadas por nematoides, foi possível avaliar quais variáveis poderiam dar maior contribuição a construção dos modelos. A partir das variáveis subdivididas em 4 épocas observamos que as variáveis mais importantes variaram de uma época para outra no decorrer do ciclo da cultura.

Através da classificação de importância pelo índice Gini para os dados da época 1, o índice CARI calculado na cena do dia 3/10/2018 mostrou menor índice de confusão nos resultados para evidenciar ocorrência de agrupamentos de plantas infectadas pela espécie de *P. brachyurus* (Fig. 5a) devido ele estar sempre nas primeiras posições das árvores de decisão geradas. A reflectância do dossel das plantas de soja no início da fase vegetativa pode ser afetada pela quantidade de REM refletida do solo com relação ao baixo índice de área foliar, afetando a interação entre a REM absorvida com a transmitida capturada pelos sensores orbitais (Curran, 1985; Feng et al., 2016). Portanto, índices de vegetação derivados de comprimentos de onda do NIR apresentaram contribuição menor nesse estágio fenológico da cultura de soja para separação entre plantas com e sem nematoides. A exatidão do modelo da época 1 foi inferior quando comparada as das demais épocas, porque a soja estava no estágio inicial vegetativo (Tab. 5). Raza et al. (2020) encontraram uma exatidão semelhante de 77% na classificação do modelo treinado pelo algoritmo RF para detecção de áreas com síndrome da morte súbita da soja no seu estágio vegetativo inicial.

Para o modelo da época 2 a classificação de importância das variáveis pelo índice de Gini mostrou que a cena do dia 7/11/2018 adicionada ao modelo para a predição de índices de vegetação não apresentou maior pureza nos resultados. A data e índice de vegetação classificados com maior relevância permaneceu sendo 3/10/2018 e índice CARI (Fig. 5b). O nematoide da espécie *P. brachyurus* atacar diretamente o sistema

radicular, fazendo com que a planta perca a alta capacidade de absorção de água e nutrientes via xilema. A deficiência na absorção de água e nutrientes se torna responsável pela degradação dos pigmentos responsáveis pela grande parte de absorção da energia no comprimento de onda do vermelho, assim, menos energia desse comprimento será absorvido, resultando em maior energia refletida (Machado, 2015). Alguns trabalhos também mostram que plantas doentes refletem mais em comprimentos de onda do vermelho (Mahlein et al., 2012; Martins et al., 2017; Bajwa et al., 2017).

O conjunto de dados que forma a classificação para época 3, obteve 9 índices de vegetação calculados em 4 datas com maior importância para o modelo (Fig. 5c). Os índices indicados a maior contribuição para a construção do modelo são derivados de bandas estreitas no comprimento espectral do vermelho, NIR e do RedEdge, esses índices de vegetação possuem boa relação na transição do estágio vegetativo para o reprodutivo da planta. A adição de dados em que o estágio fenológico da soja estava mais avançado, em aproximadamente R1 (início da florescência) influenciou para que obtivesse melhor exatidão dos agrupamentos de plantas infectadas (Tab. 5), devido a incidência e severidade da doença estarem maiores, menor interferência do solo do que no vegetativo, maior quantidade de biomassa, caracterizando melhor as diferenças espectrais entre plantas saudáveis e plantas atacadas por nematoides. (Steven et al., 1990).

Os índices de vegetação de bandas estreitas podem avaliar mudanças na vegetação em diversas escalas, além de estimar os danos causados por patógenos, pela relação que existe entre a absorvância, transmitância e reflectância nesses comprimentos de onda (Sankaran et al, 2010). Uma alta reflectância do NIR está relacionada diretamente com a estrutura interna das folhas que são responsáveis pelo processo de espalhamento inter e intra foliar (Nobel, 1999). Comprimento de onda da faixa do infravermelho próximo ou da borda do vermelho apresenta bons resultados para detecção dos sintomas causados por nematoides. (Mahlein et al., 2012; Martins & Galo, 2015;).

As variáveis mais importantes para o modelo de classificação referente à época 4 por Importância de Gini, foram índice GLI, CCCI, MCARI_MTVI2, ARVI2 e CARI2 (Fig. 5d). O índice GLI classificado como o mais importante é derivado de uma composição de bandas do visível (RGB), isso mostra que em culturas com elevado estágio de desenvolvimento, sensores RGB podem ser úteis na identificação de nematoides, o CCCI utiliza de bandas no comprimento de onda do NIR, RedEdge e na banda do vermelho, MCARI_MTVI2 com bandas específicas no comprimento de onda do vermelho, verde e NIR. Em relação a reflectância das folhas na região do NIR não tem

alteração significativa na fase final do ciclo da soja, pois há redução da área fotossintética ativa na planta (Neiverth et al., 2013). A presença de nematoides no sistema radicular das plantas podem causar alterações biofísicas e bioquímicas como: redução de biomassa e pigmentos fotossintéticos, e leva a alteração significativa na resposta espectral das plantas nos comprimentos de onda vermelho (Silva & Pereira, 2003; Jensen, 2009; Barton, 2012). Contudo, ocorre diminuição dos pigmentos da planta, gerando aumento da reflectância em comprimentos de onda relativos ao azul e ao vermelho, o índice GLI mostrou como mais importante para a classificação utilizando essas datas (Curran, 1985).

Assim, para a classificação da época 4 se utilizou imagens de todo o ciclo da cultura, mostrando o potencial no monitoramento multitemporal, principalmente no que se refere a mudanças espectrais. As plantas mais vigorosas, maiores e área com menos solo exposto permitiram diferenciar melhor plantas saudáveis de plantas doentes. O aumento da exatidão mostrou que imagens do estágio reprodutivo até o final do ciclo da cultura da soja foram importantes para a predição de locais com agrupamentos de plantas que apresentavam sintomas de nematoides (Tab. 5). O estabelecimento de uma sistematização para classificação de imagens de satélites, bem como a escolha da técnica de classificação mais adequada ao estudo pretendido, influencia diretamente na qualidade do produto, assim, como na exatidão das análises a serem realizadas. Dessa forma, a técnica de classificação (Random Forest) utilizada apresentou resultado satisfatório para a tarefa de classificação multitemporal da área de estudo, oferecendo confiabilidade para futuros estudos sobre a dinâmica de identificação de reboleiras de nematoides utilizando imagens de satélites.

5. Conclusões

O processo de classificação de imagens de satélites apresentou resultados suficientes para a avaliação multitemporal das respostas espectrais de agrupamentos de plantas de soja infestadas por *P. brachyurus*. Os modelos de classificação para as épocas 1 e 2 foram melhores com uso de índices de vegetação que utilizam comprimentos de onda do visível. Para época 3 o modelo para estimativa de *P. brachyurus* foi melhor ajustado com índices derivados de bandas e comprimentos de ondas estreitas que atuam na faixa do visível e do infravermelho. O modelo classificador da época 4 indicou que imagens referentes ao final de ciclo interferem na classificação e mostram novamente que as bandas espectrais na faixa do visível foram melhores para detectar *P. brachyurus* em cultura de soja.

6. Referências Bibliográficas

- Abreu, K.M.P. de, Coutinho, L.M., 2014. Remote sensing applied to the study of vegetation with emphasis on index of vegetation and landscapes metrics. *Revista Vértices* 16, 173–198.
- Ali, M.A., Abbas, A., Azeem, F., Javed, N., Bohlmann, H., 2015. Plant-nematode interactions: From genomics to metabolomics. *International Journal of Agriculture and Biology* 17, 1071–1082.
- Alves, T.M., Macrae, I. V., Koch, R.L., 2015. Soybean Aphid (Hemiptera: Aphididae) Affects Soybean Spectral Reflectance. *Journal of Economic Entomology* 108, 2655–2664.
- Alves, T. M., Monn, R. D., MacRae, I. V., Kock, R. L., 2019. Optimizing band selection for spectral detection of *Aphis glycines* Matsumurain soybean. *Pest Manag Sci.* 75, 942-949.
- Bajwa, S.G., Rupe, J.C., Mason, J., 2017. Soybean disease monitoring with leaf reflectance. *Remote Sensing*, 9, 1–14.
- Barton, C.V. M., 2012. Advances in remote sensing of plant stress. *Plant and Soil* 354, 41–44.
- Blevins, D. G., Dropkin, V. H., Luedders, V. D., 1995. Macronutrient uptake, translocation, and tissue concentration of soybeans infested with the soybean cyst nematode and elemental composition of cysts isolated from roots. *Journal of plant nutrition*, 18, 579-591.
- Breiman, L., 2001. Random forests. *Machine learning*, 45, 5-32.
- Breiman, L., Cutler, A., 2003. Setting up, using, and understanding random forests 4. University of California, Department of Statistics.
- Campos, J., Llop, J., Gallart, M., García-Ruiz, F., Gras, A., Salcedo, R., Gil, E., 2019. Development of canopy vigour maps using UAV for site-specific management during vineyard spraying process. *Precision Agriculture* 20, 1136–1156.
- Cao, F., Liu, F., Guo, H., Kong, W., Zhang, C., He, Y., 2018. Fast detection of sclerotinia sclerotiorum on oilseed rape leaves using low-altitude remote sensing technology. *Sensors (Switzerland)* 18.
- Chan, J.C.W., Paelinckx, D., 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment* 112, 2999–3011.
- Chemura, A., Mutanga, O., Dube, T., 2017. Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions. *Precision Agriculture* 18, 859–881.

- CONAB - Companhia Nacional de Abastecimento. 2020. Boletim de Grãos setembro 2020. <https://www.conab.gov.br/info-agro/safras/graos> > (Acessado em outubro 2020).
- Congalton, R. G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37, 35–46.
- Congalton, R. G., Green, K., 2019. Assessing the accuracy of remotely sensed data: principles and practices. CRC press.
- Costa Manso, E. S. B. G., Tenente, R. C. V., Ferraz, L. C. C. B., Oliveira, R. S., Mesquita, R., 1994. Catálogo de nematoides fitoparasitas encontrados associados a diferentes tipos de plantas no Brasil. Brasília: EMBRAPA.
- Curran, P.J., 1985. Princípios de sensoriamento remoto. Longman Inc.
- Datt, B., 1999. A new reflectance index for remote sensing of chlorophyll content in higher plants: Tests using Eucalyptus leaves. *Journal of Plant Physiology* 154, 30–36.
- El-Shikha, M. D., Barnes, E. M., Clarke, T. R., Hunsaker, D. J., Haberland, J. A., Pinter Jr, P. J., Waller, P. M., Thompson, T. L., 2008. Remote sensing of cotton nitrogen status using the Canopy chlorophyll content index (CCCI). *Transactions of the ASABE*, 51, 73-82.
- Feng, W., Zhanga, H. Y., Zhanga, Y.S., Qi, S.L., Heng, Y.R., Guo, B. B., Ma, D.Y., Guo, T. C., 2016. Remote detection of canopy leaf nitrogen concentration in winter wheat by using water resistance vegetation indices from in-situ hyperspectral data. *Field Crops Research* 198, 238–246.
- FIESP - Federação das Indústrias do estado de São Paulo. Safra mundial de soja. <<https://www.fiesp.com.br/indices-pesquisas-e-publicacoes/safra-mundial-de-soja/>>. (Acessado em abril 2020).
- Forge, T.A., DeYoung, R., Vrain, T.C., 1998. Temporal changes in the vertical distribution of *Pratylenchus penetrans* under raspberry. *Journal of Nematology* 30, 179–183.
- Franchini, J.C., Debiasi, H., Dias, W.P., Ramos Jr, E.U., Silva, J.F.V., 2014. Perda de produtividade da soja em área infestada por nematoide das lesões radiculares na região médio norte do Mato Grosso. *Agricultura de Exatidão: Resultados De Um Novo Olhar*. 274–278.
- Ghimire, B., Rogan, J., Rodríguez-Galiano, V., Panday, P., Neeti, N., 2012. An Evaluation of Bagging, Boosting, and Random Forests for Land-Cover Classification in Cape Cod, Massachusetts, USA, *GIScience & Remote Sensing*, 49, 623-643.
- Gitelson, A.A., Keydan, G.P., Merzlyak, M.N., 2006. Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters* 33, 2–6.

- Gornitz, V., 1991. Global coastal hazards from future sea level rise. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 89, 379-398.
- Huete, A., Didan, K., Leeuwen, W.J., Jacobson, A., Solanos, R., Laing, T.D., 1999. MODIS vegetation index (MOD 13): algorithm theoretical basis document. Washington: National Aeronautics and Space Administration.
- Jardim, I.N., Oliveira, D.F., Silva, G.H., Campos, V.P., de Souza, P.E., 2018. (E) - cinnamaldehyde from the essential oil of *Cinnamomum cassia* controls *Meloidogyne incognita* in soybean plants. *Journal of Pest Science* 91, 479–487.
- Jensen, J. R., 2011. *Sensoriamento Remoto do Ambiente*. 2.ed. São José dos Campos, Parêntese.
- Joalland, S., Screpanti, C., Varella, H.V., Reuther, M., Schwind, M., Lang, C., Walter, A., Liebisch, F., 2018. Aerial and ground based sensing of tolerance to beet cyst nematode in sugar beet. *Remote Sensing* 10, 1–21.
- Kaufman, Y. J., Tanre, D., 1992. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE transactions on Geoscience and Remote Sensing*, 30, 261-270.
- Kim, M. S., Daughtry, C. S. T., Chappelle, E. W., McMurtrey, J. E., Walthall, C. L., 1994. The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation (A_{par}).
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Benesty, M., 2017. *Caret: classification and regression training*. R package version, 4.
- Landis, J., Koch, G., 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33, 159-174.
- Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: A review. *Sensors (Switzerland)* 18, 1–29.
- Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. *R News* 2 3, 18–22.
- Lobo, J.M., Tognelli, M.F., 2011. Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. *Journal for Nature Conservation* 19, 1–7.
- Louhaichi, M., Borman, M. M., & Johnson, D. E., 2001. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto International*, 16, 65-70.
- Machado, A. C. Z., 2015. Nematoides devastam lavouras de soja. *Revista Campo & Negócio*, Uberlândia. <http://www.revistacampoenegocios.com.br/nematoidesdevastam-lavouras-de-soja>. (Acessado em novembro 2019).

- Mahlein, A.K., Oerke, E.C., Steiner, U., Dehne, H.W., 2012. Recent advances in sensing plant diseases for precision crop protection. *European Journal of Plant Pathology* 133, 197–209.
- Marin, D.B., de Carvalho Alves, M., Pozza, E.A., Belan, L.L., de Oliveira Freitas, M.L., 2019. Multispectral radiometric monitoring of bacterial blight of coffee. *Precision Agriculture* 20, 959–982.
- Martins, G.D., Trindade Galo, M.D.L.B., 2015. In situ spectral characterization of sugar cane infected by nematodes and *Migdolus fryanus*. *Boletim de Ciencias Geodesicas* 21, 783–796.
- Martins, GD, Galo, M. D. L. B. T., Vieira, B. S., 2017. Detecção e mapeamento da infecção do nematóide das galhas na cultura do café usando medições de sensoriamento remoto. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10, 5395-5403.
- Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing* 39, 2784–2817.
- Neiverth, W., Crusiol, L.G.T., Carvalho, J.F.C., Ferreira, L.C., Rio, A., Sibaldelli, R.N.R., Neumaier, N., Farias, J.R.B., 2013. NDVI de estádios fenológicos da soja BRS 284 de crescimento indeterminado em casa de vegetação. VIII Jornada Acadêmica da Embrapa Soja 6.
- Nobel, P. S., 1999. *Physicochemical and environmental plant physiology*. London: Academic Press, 01- 474.
- Pal, M., 2005. Random forest classifier for remote sensing classification, *International Journal of Remote Sensing*, 26: 1, 217-222.
- Ponzoni, F. J., Shimabukuro, Y. E., Kuplich, T. M., 2007. *Sensoriamento remoto no estudo da vegetação*, 127. São José dos Campos: Parêntese.
- Pudasaini, M. P., Schomaker, C. H., Been, T. H., & Moens, M., 2006. Vertical distribution of the plant-parasitic nematode, *Pratylenchus penetrans*, under four field crops. *Phytopathology*, 96, 226-233.
- Raza, M.M., Harding, C., Liebman, M., Leandro, L.F., 2020. Exploring the potential of high-resolution satellite imagery for the detection of soybean sudden death syndrome. *Remote Sensing* 12.
- Riedell, W. E., Blackmer, T. M., 1999. Leaf reflectance spectra of cereal aphid-damaged wheat. *Crop Science*, 39, 1835-1840.
- Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 67, 93–104.

- Sankaran, S., Mishra, A., Ehsani, R., Davis, C., 2010. A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture* 72, 1-13.
- SBN - Sociedade Brasileira de Nematologia. 2017. Nematoides causam prejuízos de R\$ bilhões ao ano. < <https://nematologia.com.br/index.php?page=informativo> >. (Acessado em dezembro 2019).
- Shi, D., Yang, X., 2016. An assessment of algorithmic parameters affecting image classification accuracy by random forests. *Photogrammetric Engineering and Remote Sensing* 82, 407-417.
- Shiratsuchi, L. S., Christoffoleti, P. J., Fontes, J. R. A., 2003. Aplicação localizada de herbicidas. Embrapa Cerrados-Documents (INFOTECA-E).
- Silva, R.A., Pereira, L.C., 2003. Efeitos de densidades populacionais de *Pratylenchus brachyurus* na produtividade de duas cultivares de soja, em condições de campo. *Nematologia Brasileira*, 27-268.
- Steven, M. D., Malthus, T. J., Demetriades-Shah, T. H., Danson, F. M., Clark, J. A., 1990. High spectral resolution indices for crop stress. In.: Steven MD, Clark JA Applications of remote sensing in agriculture. London: Butterworths, 209-228.
- Taquia, L. A. A., 2015. Descripción, Descarga y Manipulación de Productos Sentinel-2.
- Vale, J.R.B., da Costa, A.J., dos Santos, J.F., da Silva, E.L.S., Favacho, A.T., 2018. Análise Comparativa De Métodos De Classificação Supervisionada Aplicada Ao Mapeamento Da Cobertura Do Solo No Município De Medicilândia, Pará. *InterEspaço: Revista de Geografia e Interdisciplinaridade* 4, 26.
- VanDerWal, J., Shoo, L.P., Graham, C., Williams, S.E., 2009. Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling* 220, 589-594.
- Xie, G.H., Cui, H.D., Dong, Y., Wang, X.Q., Li, X.F., Deng, R.K., Wang, Y., Xie, Y., 2016. Crop rotation and intercropping with marigold are effective for root-knot nematode (*Meloidogyne* sp.) control in angelica (*Angelica sinensis*) cultivation. *Canadian Journal of Plant Science* 97, 26–31.
- Yang, C., 2018. High resolution satellite imaging sensors for precision agriculture. *Frontiers of Agricultural Science and Engineering*, 5, 393-405.
- Zhang, M., Qin, Z., Liu, X., Ustin, S.L., 2003. Detection of stress in tomatoes induced by late blight disease in California, USA, using hyperspectral remote sensing. *International Journal of Applied Earth Observation and Geoinformation* 4, 295-310.
- Zheng, H., Cheng, T., Li, D., Zhou, X., Yao, X., Tian, Y., Cao, W., Zhu, Y., 2018. Evaluation of RGB, color-infrared and multispectral images acquired from unmanned aerial systems for the estimation of nitrogen accumulation in rice. *Remote Sensing* 10.