

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS DA NATUREZA
CAMPUS LAGOA DO SINO

Mateus Gomes Moreira

COMPARAÇÃO ENTRE O MÉTODO DE MAXVER E *RANDOM FOREST* NA
IDENTIFICAÇÃO DE *Pinus sp.* EM ÁREAS DE CAMPO ÚMIDO

Buri

2021

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS DA NATUREZA
CAMPUS LAGOA DO SINO

Mateus Gomes Moreira

COMPARAÇÃO ENTRE O MÉTODO DE MAXVER E *RANDOM FOREST* NA
IDENTIFICAÇÃO DE *Pinus sp.* EM ÁREAS DE CAMPO ÚMIDO

Trabalho de Conclusão de Curso apresentado
como exigência parcial para a obtenção do grau
de Bacharel em Engenharia Ambiental na
Universidade Federal de São Carlos.

Orientação: Prof. Dr. Paulo Guilherme Molin

Buri

2021

Moreira, Mateus Gomes

Comparação entre o método de MAXVER e Random Forest na identificação de Pinus sp. em áreas de campo úmido / Mateus Gomes Moreira -- 2021.
33f.

TCC (Graduação) - Universidade Federal de São Carlos, campus Lagoa do Sino, Buri

Orientador (a): Paulo Guilherme Molin

Banca Examinadora: Paulo Guilherme Molin, Catherine Torres de Almeida, Felipe Rosafa Gavioli

Bibliografia

1. Sensoriamento remoto. 2. Maxver e random forest. 3. Pinus sp.. I. Moreira, Mateus Gomes. II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática (SIn)

DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Lissandra Pinhatelli de Britto - CRB/8 7539

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS DA NATUREZA
CURSO DE GRADUAÇÃO EM ENGENHARIA AMBIENTAL

Assinatura dos membros da comissão examinadora que avaliou e aprovou a Defesa de Trabalho de Conclusão de Curso do(a) candidato(a) Mateus Gomes Moreira, realizada em 07/12/2021:



Prof. Dr. Paulo Guilherme Molin – Orientador(a)
Centro de Ciências da Natureza – UFSCar – Campus Lagoa do Sino.



Dra. Catherine Torres de Almeida – Avaliadora 1
USP - ESALQ



Me. Felipe Rosafa Gavioli – Avaliador 2
CETESB e UFSCar-Sorocaba

DEDICATÓRIA

À toda minha família, em especial, aos meus pais, que tanto se dedicaram para que eu conseguisse realizar esta graduação.

AGRADECIMENTO

Como jovem, 23 anos de idade, reconhecendo a origem de meus pais, de meus avós, e entendendo um pouco a realidade de onde vieram, a realidade em que viveram, as batalhas que batalharam para ter sempre à mesa a refeição de cada dia, para que pudessem fazer crescer uma família de 8 irmãos por parte de mãe e 7 irmãos por parte de pai; entendendo que nunca cheguei perto, de fato, de conhecer a fome, de conhecer o sentimento de impotência por não poder ter o minimamente necessário para viver uma vida digna; entendendo que minha geração é economicamente muito diferente de duas gerações atrás de minha família, e que hoje, posso desfrutar da vida que possuo, bem como continuar a construí-la com maior apoio, em função do grandioso esforço que tiveram tempos atrás. Entendendo tudo isso, ainda que para muitas pessoas possa não parecer grande coisa, chegar até aqui é, sem dúvidas, uma grande conquista para mim e para toda minha família. É fruto de décadas de dedicação diária, fruto de um legado maravilhoso que ainda está sendo construído; é fruto de pais unidos e presentes, sobretudo de uma família unida sob um amor divino, e que possui orgulho em se dedicar para que a geração seguinte seja melhor que a anterior.

Isso tudo descrito, agradeço:

Primeiramente, à Deus, pela vida, por todo amor e carinho, por todas as oportunidades, por toda força e guia, pela minha família e amigos e por todas as pessoas que colocou em meu caminho.

À minha família, que tudo que pôde fez para que eu conseguisse vencer cada desafio, que tanto amor e carinho me deu para que eu não desanimasse da caminhada.

À cada um de meus professores, que tanto se esforçaram para formar nesta pessoa o melhor profissional. Em especial, ao Professor Doutor Paulo Guilherme Molin, por aceitar ser meu orientador e levar a função com dedicação e gentileza, e porque se tornou para mim uma referência de sucesso profissional e pessoal; ao Professor Doutor André M. A. Toledo, pela parceria e pelo apoio; e ao Professor Doutor Naaman, pelos ensinamentos durante a liderança da empresa júnior. Entre outros tantos professores que eu gostaria de citar aqui. A cada um de vocês, reforço: meu muito obrigado!

À LS Consultoria Júnior, que me permitiu vivenciar, sem dúvidas, muitas das melhores experiências que tive na vida e não apenas na graduação, bem como adquirir muitos

aprendizados; e ao CePE-Geo, que me permitiu conhecer melhor a pesquisa científica na prática e abriu as portas para que eu pudesse realizar as pesquisas para este trabalho.

Aos meus tão queridos amigos e colegas de graduação, os quais foram essenciais em minha formação de vida, em quem me tornei hoje. Em especial, ao Felipe Tassi, que dividiu comigo a mesma casa desde as primeiras semanas da graduação, que me acompanhou de perto nas minhas tantas dificuldades iniciais, e que muito me ajudou no desenvolvimento de meu aprendizado, pelas aventuras que vivemos (saiba você, meu querido, que você foi muito importante para que eu conseguisse chegar até aqui – Muito obrigado!); Ao Mario Gabriel e à Bianca Jacobo, que também dividiram casa comigo e acompanharam minha jornada de modo tão próximos; Ao Matheus Mônaco, Vinícius Orlandi, Vinícius Rainer, Teresa Xavier (obrigado por toda parceria, velha! Junto com a Júlia, você foi fundamental para me manter ativo na EJ) e Júlia Andrade (junto com a Teresa, você foi essencial para seguirmos firmes na EJ – obrigado, Juh!), Gabriela Ribeiro, Alexandre do Carmo, Bruna Renata, Christian Price e à tantas outras pessoas que considero muito! Agradeço ainda, em especial, à Bianca Albuquerque, que se tornou uma pessoa muitíssimo importante em minha vida: obrigado, Bia, por todo amor e carinho!

E a todos aqueles com quem convivi ao longo destes anos todos de graduação, que, direta ou indiretamente, contribuíram para minha formação. A cada um vocês aqui citados, meu muito obrigado! Desejo muito sucesso na caminhada de cada um!

Um grande abraço!

“Quem não vive para servir, não serve para viver.”

(Mahatma Gandhi)

RESUMO

MOREIRA, Mateus Gomes. **Comparação entre o método de MAXVER e *Random Forest* na identificação de *Pinus sp.* em áreas de campo úmido.** 2021. Trabalho de Conclusão de Curso – Universidade Federal de São Carlos, *campus* Lagoa do Sino, Buri, 2021.

No Brasil, o *Pinus sp.* é uma planta invasora muito comum, uma vez que apresenta excelente adaptação às condições edafoclimáticas das regiões Sudeste e Sul do país, com destaque para áreas de campo úmido. Sendo uma planta invasora, ela pode implicar na exclusão de espécies nativas através da competição, ocasionando em perda de biodiversidade e até modificações na estrutura do ecossistema. Logo, é importante dispor de ferramentas que auxiliem no planejamento de mitigação de impactos ambientais causados por estas espécies. Entre estas ferramentas está o Sensoriamento Remoto, o qual pode identificar e localizar indivíduos de plantas exóticas e invasoras. Nesse contexto, o presente trabalho teve por objetivo utilizar imagens multiespectrais de altíssima resolução para testar e comparar os classificadores MAXVER e *Random Forest* (RF) através de uma matriz de confusão, mais especificamente, o Índice Kappa, a fim de se compreender qual possui melhor desempenho na identificação de indivíduos isolados de *Pinus sp.* em uma área de campo úmido. Para isso, foram criadas 5 classes, “Pinus” (indicando indivíduos de *Pinus*), “Nao_Pinus” (indicando todos os componentes restantes da paisagem que não se enquadram nas demais classes), “Água” (indicando locais com água, como lagos, por exemplo), “Solo” (indicando solo exposto) e “Sombra” (indicando locais com sombra), com os quais executou-se os dois classificadores utilizando uma composição de imagens de 4 bandas (NIR, *Red Edge*, *Red* e *Green*). Com isso, criou-se 500 pontos aleatórios divididos igualmente por cada classe carregando a informação de sua classificação realizada pelos classificadores e a realizada manualmente, esta tomada como real. A partir destas informações, construiu-se então a Matriz de Confusão, obtendo-se o Índice Kappa. Ambos foram classificados como com concordância moderada, sendo MAXVER com índice de 0,5023 e RF com índice de 0,5264. Observando estes valores, percebeu-se que RF apresentou maior concordância que MAXVER e, portanto, dentre estes dois classificadores, sugere-se que seja melhor para a finalidade apresentada, diante da metodologia utilizada. Destaca-se ainda que, além de MAXVER e *Random Forest*, outros classificadores podem ser utilizados para identificação de indivíduos de *Pinus sp.* e outras espécies, os quais podem inclusive apresentar melhor desempenho para esta finalidade.

Palavras-chave: *Pinus sp.* Plantas invasoras. Campo úmido. Sensoriamento remoto. Imagens multiespectrais. MAXVER. *Random Forest*. Índice Kappa.

ABSTRACT

MOREIRA, Mateus Gomes. **Comparison between the MAXVER and Random Forest method to identify *Pinus sp.* in wet field areas.** 2021. Completion of Course Work – Federal University of São Carlos, *campus* Lagoa do Sino, Buri, 2021.

In Brazil, *Pinus sp.* is a very common invasive plant and has an excellent adaptation to edaphoclimatic conditions in the Southeast and South regions of the country, especially in wet field areas. As an invasive plant, it can cause the exclusion of native species through competition, causing a loss of biodiversity and even changes in the structure of the ecosystem. Therefore, having tools that help in planning the mitigation of environmental impacts caused by these species is very important. Among these tools is Remote Sensing, which can identify and locate exotic and invasive plant individuals. Thus, the present work aimed to use very high resolution multispectral images to test and compare the MAXVER and Random Forest (RF) classifiers through a confusion matrix, more specifically, the Kappa Index, to understand which one has the best performance in the identification of isolated individuals of *Pinus sp.* in wet field areas. To do that, 5 classes were created, "Pinus" (*Pinus sp.*), "Nao_Pinus" (indicating all the remaining components of the landscape that do not fit into the other classes), "Agua" (indicating places with water, such as lakes, for example), "Solo" (indicating exposed soil) and "Sombra" (indicating shady locations), with which the two classifiers were performed using a 4-band image composition (NIR, Red Edge, Red and Green). Then, 500 random points were created, divided equally by each class, carrying the information of its classification performed by the classifiers and the one performed manually, this one taken as real. From this information, the Confusion Matrix was constructed, obtaining the Kappa Index. Both were classified as having moderate agreement, being MAXVER with an index of 0.5023 and RF with an index of 0.5264. Observing these values, noticed that RF showed greater agreement than MAXVER and, consequently, among these two classifiers, RF is suggested being better for the presented purpose, under the methodology used. It is noteworthy that, in addition to MAXVER and Random Forest, other classifiers can be used to identify individuals of *Pinus sp.* and other species, which may even show better performance for this purpose.

Keywords: *Pinus sp.* Invasive plants. Wet field. Remote sensing. Multispectral images. MAXVER. Random Forest. Kappa Index.

LISTA DE FIGURAS

Figura 1 – Vista aérea da área de estudo	19
Figura 2 – Base fixada em campo	20
Figura 3 – <i>Rover</i> em uso sob um indivíduo de <i>Pinus</i>	22
Figura 4 – <i>Pinus</i> em detalhe sob pontos em vermelho em uma imagem RGB	23
Figura 5 – Área de estudo com pontos de amostra.....	24
Figura 6 - Classificação pelo método de MAXVER	26
Figura 7 - Classificação pelo método de <i>Random Forest</i>	27
Figura 8 – <i>Zoom</i> aplicado aos ortomosaicos em uma mesma região com pontos (em azul) sobre indivíduos de <i>Pinus</i> – MAXVER à esquerda e RF à direita	27
Figura 9 – <i>Zoom</i> aplicado sobre um dos lagos da área de estudo – MAXVER à esquerda e RF à direita	28

LISTA DE ABREVIATURAS E SIGLAS

GNSS – Global Navigation Satellite System

GPS – Global Positioning System

IBGE – Instituto Brasileiro de Geografia e Estatística

MAXVER – Máxima Verossimilhança

NIR – Near Infrared

RF – Random Forest

RGB – Red, Green and Blue

RTK – Real Time Kinematic

SUMÁRIO

1. INTRODUÇÃO COM FUNDAMENTAÇÃO TEÓRICA	14
1.1. SENSORIAMENTO REMOTO E MÉTODOS DE CLASSIFICAÇÃO	15
1.1.1. Método de classificação MAXVER	16
1.1.2. Método de classificação <i>Random Forest</i>	16
1. 2. OBJETIVOS	17
2. METODOLOGIA.....	17
3. RESULTADOS E DISCUSSÃO	26
4. CONCLUSÕES.....	31
REFERÊNCIAS	32

1. INTRODUÇÃO COM FUNDAMENTAÇÃO TEÓRICA

De acordo com Matos e Pivello (2009), são muito comuns impactos em um ecossistema causados pela presença das chamadas plantas invasoras. Estas plantas apresentam, na grande maioria das vezes, crescimento e reprodução rápidos e intensos nas comunidades onde se encontram, podendo cobrir grandes áreas ou atingir amplas distâncias de sua origem. Suas definições variam em função do contexto: estas plantas podem ser as “pragas” ou “ervas-daninhas” (popularmente conhecidas) para os agricultores – em que num sentido ecológico, são consideradas como “colonizadoras” ou “pioneiras”; e, adicionando-se a questão biogeográfica, podem também ser espécies nativas (provenientes da própria comunidade) ou exóticas (provenientes de outros ambientes) (JAMES *et al.*, 1991; REJMANEK, 1995; CRONK e FULLER, 1995; WILLIAMSON, 1996, *apud* MATOS e PIVELLO, 2009). Os impactos causados em um ecossistema pela inserção de um indivíduo considerado invasor, segundo os estudos de Mooney e Hobbs (2000), podem levar à exclusão de espécies nativas por competição com as invasoras, ocasionando extinções locais, perda de biodiversidade e, além disso, modificações na estrutura do ecossistema, bem como sua completa descaracterização.

No Brasil, uma planta muito comum de ser observada como invasora é o *Pinus sp.*, a qual apresenta excelente adaptação às condições edafoclimáticas das regiões Sudeste e Sul do país (AMARAL *et al.*, 2015; VASQUES *et al.*, 2007). Sendo matéria-prima fundamental para movimentação do setor madeireiro do país, o qual é de grande importância para a economia brasileira, as florestas de *Pinus sp.* correspondem a uma área de aproximadamente 5 milhões de hectares, o que corresponde a 0,9% de toda a cobertura florestal nacional, sendo que destes, 57,6% está situado na região Sul do país, ou seja, mais da metade de toda cobertura de *Pinus sp.* está situado em apenas um conjunto de 3 estados dentre os 26 (SIQUEIRA, 2003 *apud* VASQUES *et al.*, 2007).

Dadas as suas características de ótima aptidão, rápido crescimento e desenvolvimento, bem como a intensa prática de silvicultura em função de suas fortes vantagens econômicas, o *Pinus* tem sido natural e facilmente espalhado para locais onde é considerado invasor (VASQUES *et al.*, 2007 e CONCIANI, 2018). Por conta de sua notável capacidade de desenvolvimento em ambientes úmidos, os denominados campos úmidos são locais frequentemente invadidos por esta planta (CONCIANI, 2018). Campos úmidos são:

[...] formações essencialmente herbáceo-subarbustivas que ocorrem em terrenos úmidos ou alagadiços em regiões savânicas. Podem ser encontrados formando

estreitas faixas de transição entre o cerrado [...] e as florestas ribeirinhas, ao longo de cursos d'água permanentes ou temporários, ocupando amplas planícies de declive suave ou associados a áreas de nascentes (TANNUS, 2007, p. 9).

Nestas condições, conhecendo os problemas ambientais que o *Pinus sp.* pode ocasionar, bem como os benefícios econômicos que se pode obter com sua extração, é importante dispor de meios para se identificar, mapear e quantificar seus indivíduos num certo território de interesse. Para isso, o sensoriamento remoto se apresenta como um ótimo conjunto de técnicas e ferramentas que tem sido muito empregado para atividades semelhantes a estas, apresentando bons resultados e se mostrando promissor (JOSHI, LEEUW e DUREN, 2004; AMARAL *et al.*, 2015; ASNER *et al.*, 2007).

Os autores Joshi, Leeuw, Duren (2004) utilizaram o sensoriamento remoto para estudarem o potencial de distribuição de espécies invasoras através da análise hiperespectral do dossel do território de interesse, obtendo a assinatura espectral das diferentes espécies analisadas. Outros autores como Xie, Roberts e Johnson (2008), Amaral *et al.* (2015) e Asner *et al.* (2007) também utilizaram de técnicas de sensoriamento remoto, analisando imagens hiperespectrais e multiespectrais, texturas, fenologias e outros a fim de analisar plantas invasoras, tal como o *Pinus sp.*

Existem 5 eixos principais que caracterizam os dados obtidos por sensores remotos: a extensão espacial (tamanho da imagem), a resolução espacial (tamanho do pixel), a resolução espectral (número e faixa de valores visíveis e de bandas de infravermelho), a resolução temporal (frequência de aquisição de dados) e a resolução radiométrica (nível de sensibilidade de um sensor para detectar pequenas variações radiométricas) (BRADLEY, 2014). Atualmente, não existe um sensor capaz de atingir alta cobertura espacial, espectral, temporal e radiométrica em uma ampla extensão espacial, o que implica que as abordagens realizadas por sensoriamento remoto com a tecnologia existente serão sempre limitadas às variações entre estes 5 eixos. Logo, a revisão e a busca pelo aprimoramento de técnicas de sensoriamento remoto se tornam necessárias a fim de melhorar a eficiência e eficácia de seu uso.

Vale ressaltar ainda que, diante disso, aeronaves remotamente pilotadas (RPAs), com sensores acoplados, surgem como uma plataforma que pode unir alguns dos eixos apresentados.

1.1. SENSORIAMENTO REMOTO E MÉTODOS DE CLASSIFICAÇÃO

Dispondo de imagens da área de interesse para analisar espécies invasoras, é possível realizar a classificação das imagens de diferentes formas. Assim, existem duas formas

principais de se classificar uma imagem: supervisionada, a qual se realiza o treinamento do classificador com base em informações conhecidas que permitem a identificação de cada classe de interesse; e a não supervisionada, também denominada como classificação automática, em que é usado um algoritmo para reconhecer as classes presentes na imagem (SOUZA *et al.*, 2011; GONÇALVES, 2019).

Neste contexto, para a classificação supervisionada, dois métodos de classificação existentes são os denominados MAXVER e *Random Forest*.

1.1.1. Método de classificação MAXVER

O classificador por máxima verossimilhança (MAXVER) é:

[...] um algoritmo paramétrico, que associa classes considerando pontos individuais da imagem e assume que essas classes possuem distribuição normal, a partir dos parâmetros definidos e uma amostra de pixels adquiridas *a priori*, o algoritmo computa a probabilidade estatística de um pixel desconhecido pertencer a uma ou outra classe (NERY *et al.*, 2013, p. 2).

E de acordo com Liu (2006) *apud* Nery *et al.* (2013), o MAXVER se refere a um procedimento que envolve estimativas de valores médios de cada classe e da matriz de covariância de acordo com o padrão das amostras de treinamento para classificar a imagem. Dada sua efetividade, seu limiar de aceitação é de 99,9% (CÂMARA *et al.*, 1996 *apud* NERY *et al.*, 2013). Logo, as primeiras amostras realizadas para basear a aplicação do algoritmo deve ser representativa daquilo que se tem como objetivo identificar e mapear.

1.1.2. Método de classificação *Random Forest*

O *Random Forest* (“Floresta Aleatória” em tradução livre), método proposto por Breiman (2001) *apud* Guedes e Guimarães (s/d), consiste em:

[...] um conjunto de árvores de decisão construídas no momento de treinamento do método. Para construí-las são selecionados aleatoriamente alguns dos atributos contidos dentro do vetor de características. Uma vez feito isso, calcula-se a entropia apresentada por cada atributo e aquele que possuir a maior é escolhido para separar as classes naquela posição da árvore. A saída do classificador é dada pela classe que foi retornada como resposta pela maioria das árvores pertencentes à floresta. O método apresenta aprendizado não supervisionado, ou seja, dado às instâncias de treinamento e as etiquetas que indicam a classe de cada uma, o algoritmo aprende a classificá-las sem que haja a intervenção de um usuário no processo. O método vem sendo largamente utilizado nas mais diversas áreas do aprendizado de máquina e vem apresentando ótimos resultados (GUEDES e GUIMARÃES, s/d, p. 3).

Em palavras simples, este algoritmo cria árvores decisórias e as combina para obter uma predição com maior acurácia e estabilidade. Logo, analogamente ao método de MAXVER, a primeira árvore selecionada como amostra para basear a aplicação do algoritmo deve ser representativa daquilo que se tem como objetivo identificar, tal como plantas invasoras, por exemplo.

Estes dois métodos de classificação automática apresentados podem ser utilizados para a identificação e o mapeamento de *Pinus sp.*

1. 2. OBJETIVOS

O objetivo geral do presente trabalho é a realização da comparação entre os métodos de classificação MAXVER e *Random Forest* na identificação de *Pinus* em áreas de campo úmido.

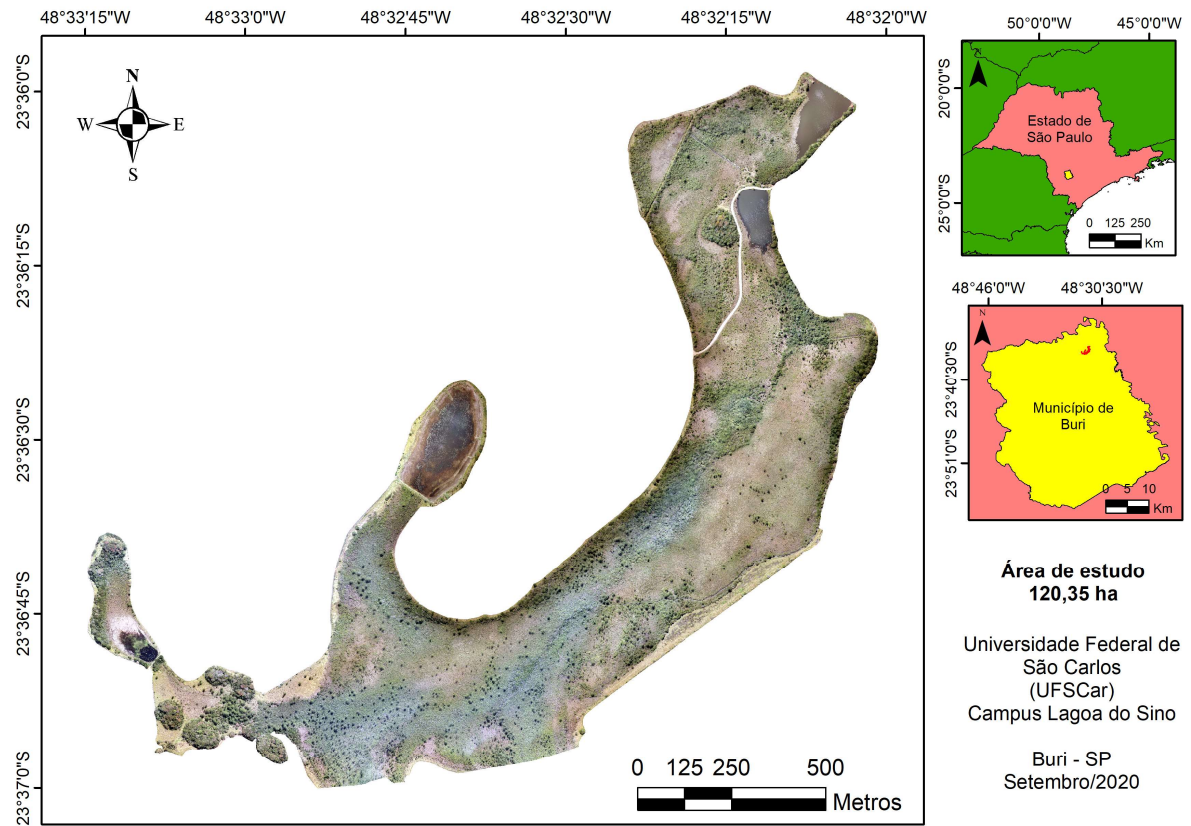
Os objetivos específicos são:

- Estudar as técnicas de análise do comportamento espectral de espécies arbóreas;
- Com equipamentos de GPS, realizar amostragens de *Pinus* isolados em áreas de campo úmido;
- Analisar a refletância do *Pinus sp.* nos pontos amostrados;
- Testar os dois métodos de classificação utilizando a refletância analisada nos *Pinus* amostrados;
- Construir a Matriz de Confusão para cada um dos classificadores e compará-las.

2. METODOLOGIA

Para realização do estudo, a área utilizada é a apresentada no Mapa 1, a qual possui 120,35 hectares compostos basicamente por lagos, estrada e, majoritariamente, campo úmido. Está localizada na fazenda do *Campus* Lagoa do Sino, da Universidade Federal de São Carlos, no município de Buri, estado de São Paulo. À sua volta, estão extensas áreas de plantio de alta produtividade, com rotação de três diferentes culturas: milho, trigo e soja. Sobre o local em si, não fora utilizado para atividades agrícolas, sendo uma área de campo úmido contendo múltiplas nascentes, o que torna o solo instável para utilização de veículos agrícolas.

Mapa 1 – Área de estudo



Fonte: Autoria própria.

A área foi escolhida justamente por contemplar campos úmidos, os quais, segundo (CONCIANI, 2018), são áreas muito propícias para a presença de *Pinus sp.* Visitando a área, é nítida a presença de grande quantidade de indivíduos da espécie, como se vê na Figura 1.

Figura 1 – Vista aérea da área de estudo



Fonte: Autoria própria.

Assim, utilizando o drone eBee X da senseFly, com tecnologia RTK e equipado com uma câmera S.O.D.A. e um sensor multiespectral Parrot Sequoia+, fez-se o sobrevoo da área em meados de setembro, em época de seca, em um dia ensolarado, com poucas nuvens e no período da tarde, por volta das 15h00. Durante o voo, usou-se uma base GNSS RTK modelo Hiper V, da Topcon, para, posteriormente em laboratório, realizar a correção das cenas obtidas pelo drone. Este equipamento permitiu obter uma precisão centimétrica das coordenadas de cada cena e, portanto, a construção de um ortomosaico com proporções e localização relativamente muito próximas da realidade. Na

Figura 2, é possível observar a base utilizada no dia do voo.

Figura 2 – Base fixada em campo



Fonte: Autoria própria.

Realizados os voos, iniciou-se o processo de correção dos dados, a começar pela correção da coordenada da base, e, para melhor organização, chamou-se esta de “Base A”. Para a correção, então, usou-se o *software* Topcon Tools e, através de uma triangulação de bases, isto é, tomando como referências três bases fixas que, quando interligadas formam um triângulo que abranja o ponto da Base A, foi feita a correção de sua localização, gerando nova coordenada, essa que foi utilizada para correção das imagens. Estas três bases fixas correspondem às estações base localizadas nos municípios de Botucatu, Ourinhos e Cananéia, todos do estado de São Paulo, e seus dados foram obtidos no site do IBGE.

Obtidas as coordenadas, iniciou-se o processo de correção das imagens. Para isso, usou-se o *software* eMotion, também da senseFly, inserindo a nova coordenada obtida com auxílio

da base e gerando novas cenas corrigidas. Em seguida, utilizou-se o *software* Pix4D para realizar o processamento das cenas e gerar ortomosaicos.

O sensor Parrot Sequoia+ é multiespectral com 4 bandas: *green*, *red*, *red edge* e NIR, enquanto a câmera S.O.D.A. possui apenas o espectro visível, isto é, RGB. Logo, das cenas obtidas pela S.O.D.A., construiu-se um ortomosaico RGB, e das cenas obtidas pela Sequoia+, um ortomosaico para cada uma das 4 bandas, possibilitando a realização de uma composição de bandas. Vale ressaltar que os pixels destes ortomosaicos possuem, aproximadamente, 14 cm² de tamanho.

Logo, obtidas as cenas corrigidas e construídos os ortomosaicos RGB e multiespectrais, foi preciso coletar pontos de amostra em locais contendo *Pinus sp.*, a fim de ter a certeza, ao observar os pontos sobre os ortomosaicos, de que nos pontos apresentados existe um indivíduo de *Pinus*, podendo assim utilizá-lo para treinamento dos classificadores, a fim de obter maior precisão. Para isso, então, foi-se a campo equipado novamente com uma base GNSS RTK, o mesmo utilizado no passo anterior, e um *rover*, o qual foi utilizado para o caminhamento. Na **Erro! Autoreferência de indicador não válida.**, é possível ver o *rover* já em uso no campo.

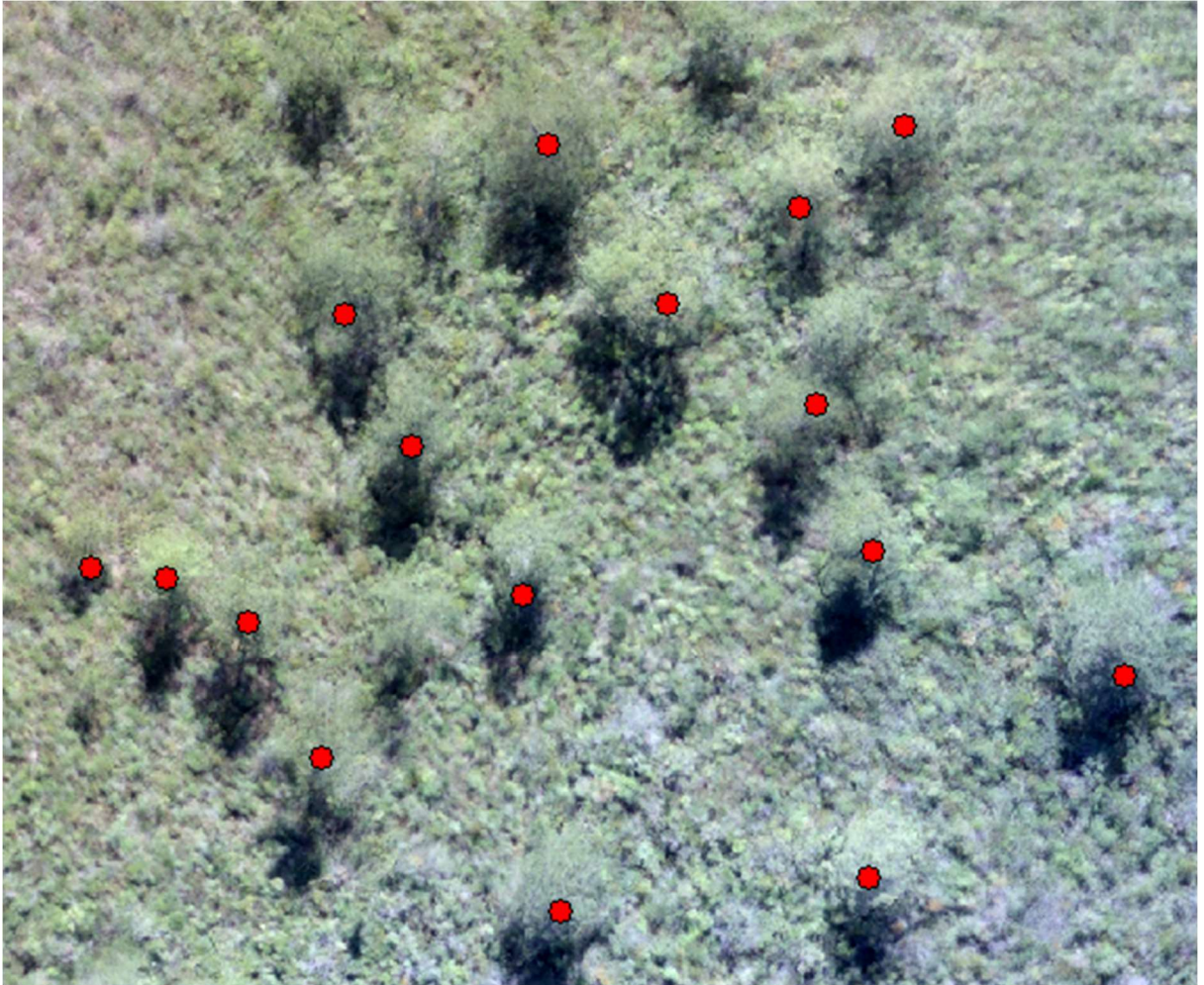
Figura 3 – Rover em uso sob um indivíduo de *Pinus*



Fonte: Autoria própria.

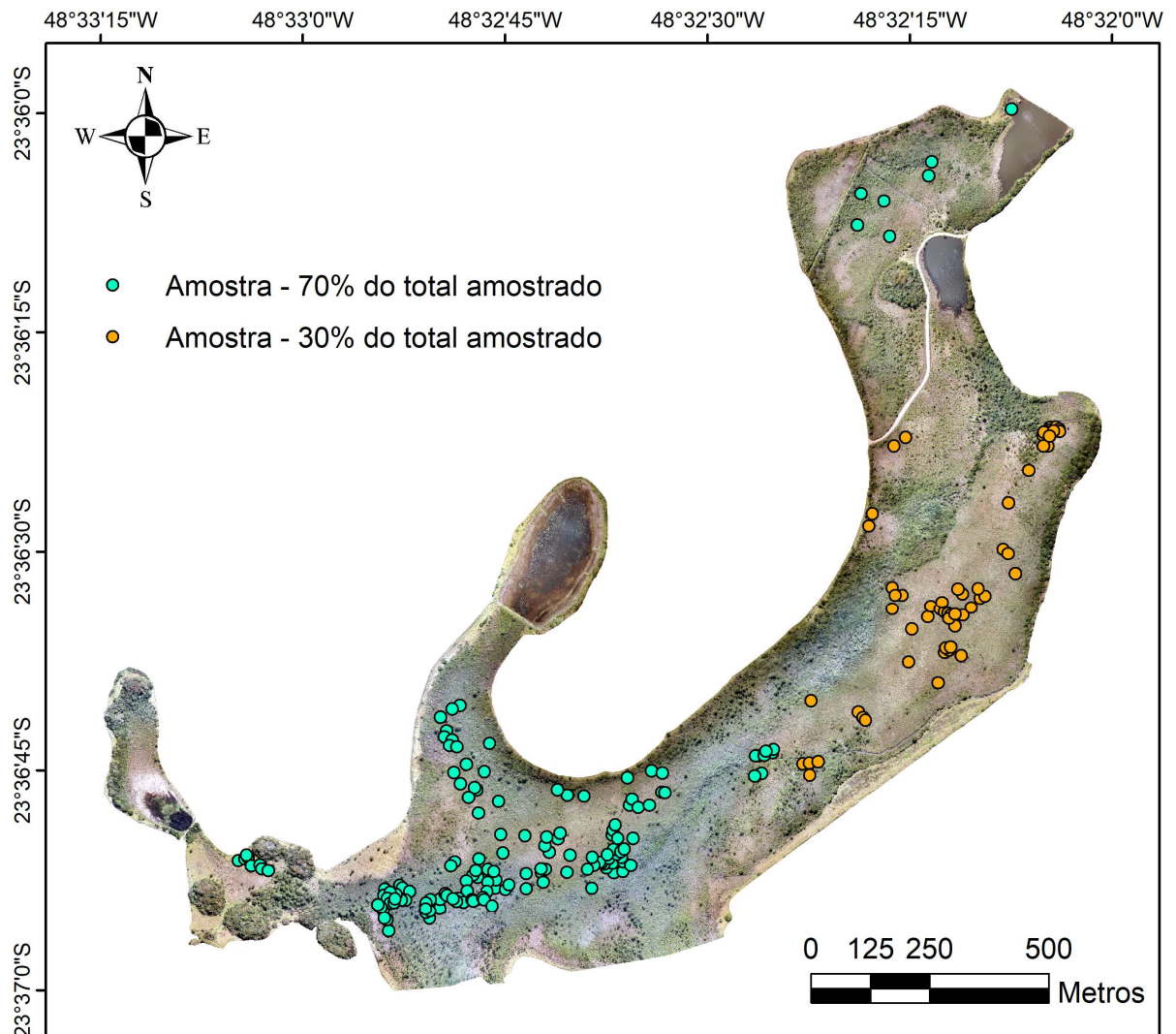
Inicialmente, tinha-se por objetivo coletar 200 pontos para amostra. Entretanto, com o andamento da atividade, percebeu-se a dificuldade do caminhamento nas áreas de campo úmido em função da densidade da vegetação, a qual inviabilizou a atividade e exigiu outra forma de identificação dos pontos referentes aos indivíduos de *Pinus*. Assim, uma vez que o ortomosaico RGB obtido com o drone possui resolução suficiente para identificação visual das árvores, passou-se a realizar a identificação dos indivíduos de *Pinus* e pontuá-los para amostra. Na Figura 4 é possível observar indivíduos de *Pinus* em maiores detalhes sob os pontos em vermelho. E na Figura 5 é possível observar a área de estudo contendo os *Pinus* amostrados através das imagens obtidas com o drone.

Figura 4 – *Pinus* em detalhe sob pontos em vermelho em uma imagem RGB



Fonte: Autoria própria.

Figura 5 – Área de estudo com pontos de amostra



Fonte: Autoria própria.

Dessa forma, amostrou-se ao todo 206 pontos, sendo destes 70% utilizados para treinamento dos classificadores, e o restante usado para validação. Entretanto, realizou-se posteriormente a validação através da criação automática de pontos aleatórios e já classificados, a partir dos quais construiu-se a matriz de confusão. Portanto, os 30% dos pontos que seriam utilizados para validação não foram utilizados.

Assim, criados os pontos e utilizando o *software* ArcGIS Pro, foi possível criar polígonos de amostra com auxílio do ortomosaico em RGB sobre os pontos estabelecidos, os quais auxiliaram no direcionamento para encontrar indivíduos de *Pinus*. Logo, criou-se, primeiramente, 2 classes: “Pinus” e “Nao_Pinus”, em que os polígonos amostrais equivalentes

à classe “Pinus” foram criados sobre a copa dos indivíduos de *Pinus* referenciados pelos pontos amostrados, enquanto os polígonos equivalentes ao “Nao_Pinus” contemplam amostras aleatórias que não são *Pinus*. Ao todo, foram 148 polígonos de amostra na classe “Pinus” e 71 polígonos na classe “Nao_Pinus”.

Dispostos das classes, foram realizadas as primeiras classificações das imagens por MAXVER e *Random Forest*. Entretanto, era nítida a confusão dos classificadores e, portanto, percebeu-se a necessidade de incluir mais classes além de “Pinus” e “Nao_Pinus”, acrescentando-se, então, as classes “Agua”, “Solo” e “Sombra”, a fim de diversificar a classificação, sendo então 5 classes ao todo. Na classe “Agua”, foram amostrados polígonos sobre os lagos; em “Solo”, foram amostrados polígonos sobre áreas nitidamente de solo exposto, como a estrada; e em “Sombra”, sobre áreas que contemplam sombras. Para cada uma destas classes foram feitos 26 polígonos de amostra e, após estas, foram refeitas as amostras para “Pinus” e “Nao_Pinus”.

Criadas, então, as novas classes e atualizadas as já existentes, rodou-se novamente os classificadores através do *software* ArcGIS Pro, gerando um *raster* para cada classificador. A fim de se compreender a área de cada classe, usou-se a ferramenta “*Raster to Polygon*” (*Raster* para Polígono – em tradução direta) para transformar o *raster* e suas diferentes classes em polígonos, os quais permitem contabilizar área.

Gerados os *rasters*, deu-se início ao processo de criação da matriz de confusão, mais especificamente o Índice Kappa. Para isso, criou-se, em cada *raster* gerado, 100 pontos aleatórios em cada uma das classes através da ferramenta “*Create Accuracy Assessment Points*” (Criar pontos de avaliação de precisão – em tradução direta). Dessa forma, para cada ponto criado havia uma associação à classe que o contemplava – exemplo: a classe “Agua” foi denominada como classe 1, logo, os pontos que, aleatoriamente, estavam sobre esta classe, recebiam em sua tabela de atributos a classificação 1. Portanto, em cada um dos 500 pontos criados, foi possível compreender em qual classe se situava de acordo com cada classificador.

Disposto da classificação de cada ponto segundo os classificadores, era necessário compreender o que cada ponto representava segundo a realidade, ao invés dos resultados do classificador através das amostras. Isso permitiria avaliar com o Índice Kappa a precisão do classificador. Logo, para cada um dos 500 pontos gerados em cada *raster*, foi realizada uma classificação manual e visual com auxílio do ortomosaico em RGB. Ou seja, observou-se o local sobre o qual cada ponto se situou e dava-lhe o número da classe que o compreendia. Ao

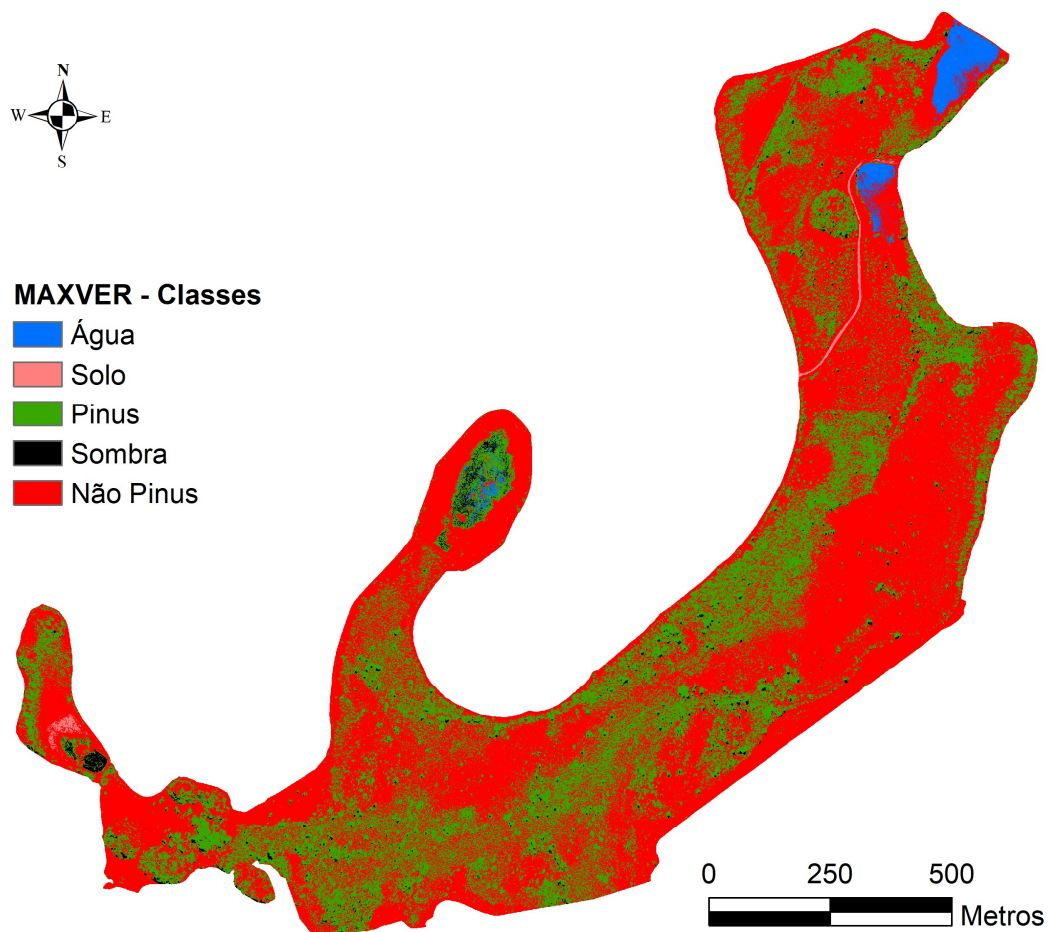
fim, a tabela de atributos dos conjuntos de pontos continha 3 colunas: uma referente à numeração do ponto, outra com as classes determinadas pelos classificadores de forma automática e outra com as classes determinadas visual e manualmente.

Disposto das tabelas de atributos, uma referente ao classificador de MAXVER e outra referente ao de *Random Forest*, calculou-se o Índice Kappa através da ferramenta “*Compute Confusion Matrix*” (Computar Matriz de Confusão – em tradução direta), obtendo, então, um valor do Índice Kappa para cada classificador estudado.

3. RESULTADOS E DISCUSSÃO

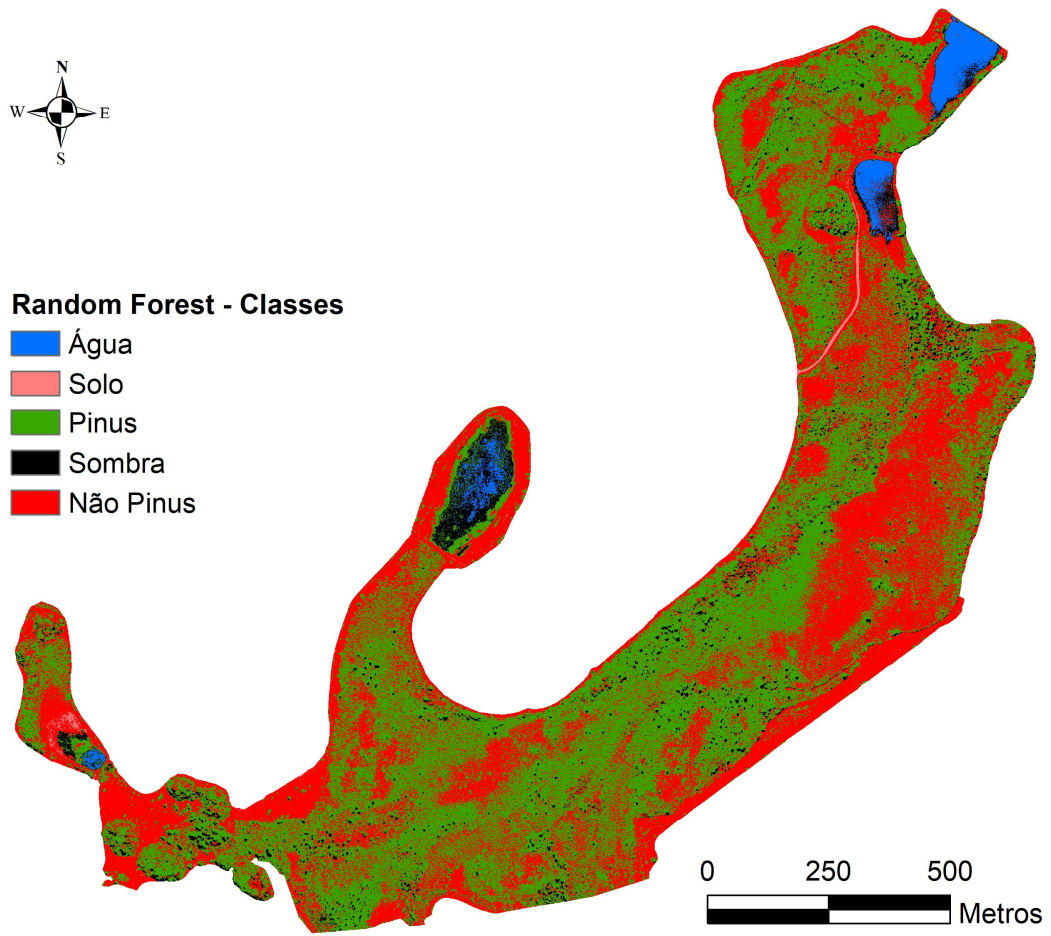
Rodados os classificadores, foram construídas as Figura 6 e 7, e, a partir delas, a Figura 8, com maiores detalhes.

Figura 6 - Classificação pelo método de MAXVER



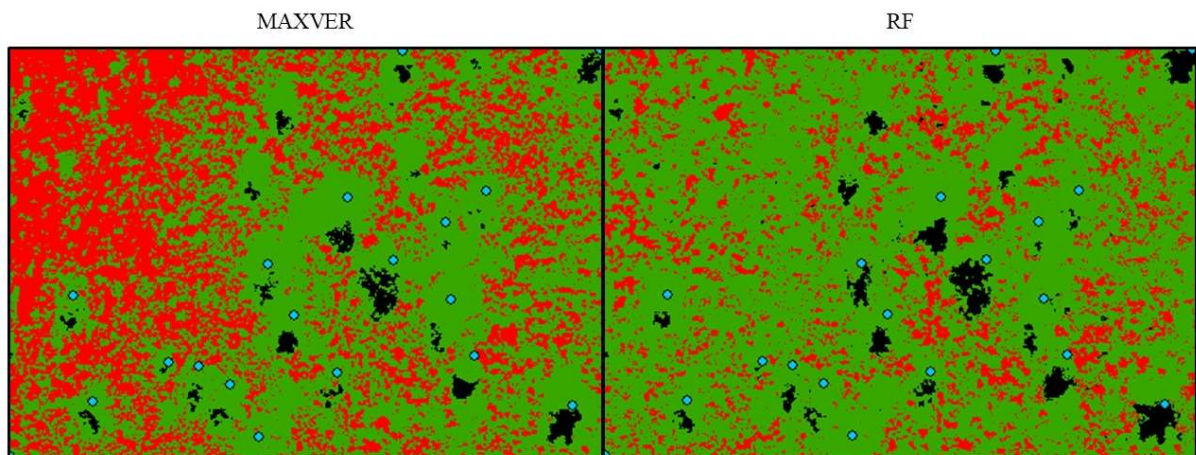
Fonte: Autoria própria.

Figura 7 - Classificação pelo método de *Random Forest*



Fonte: Autoria própria.

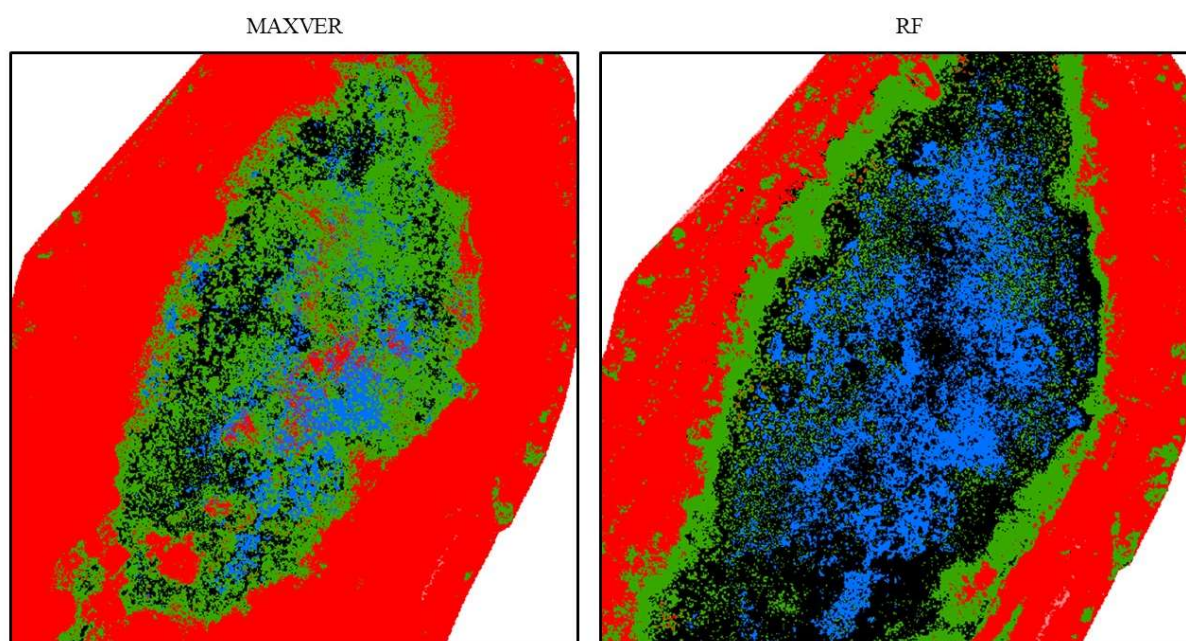
Figura 8 – *Zoom* aplicado aos ortomosaicos em uma mesma região com pontos (em azul) sobre indivíduos de *Pinus* – MAXVER à esquerda e RF à direita



Fonte: Autoria própria.

Observando as figuras, nota-se que o classificador RF estima mais *Pinus* e abrange mais áreas desta classe de forma contínua, enquanto os *Pinus* estimados por MAXVER ocupam áreas mais descontínuas e com um aspecto de pontos ligeiramente mais espalhados, o que é coerente quando se busca identificar árvores isoladas. Entretanto, por mais que não seja o foco, observando as demais classes, nota-se que ambos os classificadores indicaram “Sombra” ou “*Pinus*” em locais que deveriam ser denominados como “Água”, como é possível observar na Figura 9, e que sugere um erro grotesco no processo de classificação.

Figura 9 – Zoom aplicado sobre um dos lagos da área de estudo – MAXVER à esquerda e RF à direita



Fonte: Autoria própria.

Estes erros podem ser ocasionados em função da quantidade e qualidade das amostras utilizadas para testar os classificadores, ou ainda, pelo fato de que a metodologia e/ou os equipamentos utilizados não sejam capazes de distinguir os pixels de acordo com o objetivo apresentado.

Após a geração das figuras, transformou-se as áreas de cada uma das classes em polígonos, o que permitiu calcular o tamanho de cada uma delas. O Gráfico 1 apresenta a área obtida para cada classe em cada um dos métodos de classificação.

Tabela 2 – Índice Kappa para RF

		Random Forest							
Classificação Manual	Classe	Agua	Solo	Pinus	Sombra	Nao Pinus	Total	U Acurácia	Kappa
	Agua	10	0	0	0	0	10	1	-
	Solo	0	9	0	0	1	10	0,9	-
	Pinus	4	1	135	2	107	249	0,5422	-
	Sombra	6	0	3	13	2	24	0,5417	-
	Nao Pinus	2	6	10	0	197	215	0,9163	-
	Total	22	16	148	15	307	508	0	-
	P Acurácia	0,4545	0,5625	0,9122	0,8667	0,6417	0	0,7165	-
	Kappa	-	-	-	-	-	-	-	0,5264

De acordo com Bussab e Morettin (2017), o Índice Kappa possui as seguintes correspondências:

Tabela 3 – Valores de Kappa

Intervalo	Descrição
<0	Ausência de concordância
0-0,19	Concordância pobre
0,20-0,39	Concordância leve
0,40-0,59	Concordância moderada
0,60-0,79	Concordância substantiva
0,80-1,00	Concordância quase perfeita

Observando os índices obtidos para cada classificador e comparando-os com a Tabela 3, tem-se que ambos apresentaram concordância moderada. Entretanto, o Índice Kappa do classificador RF é maior e, portanto, sugere apresentar melhor eficácia para identificação de *Pinus*, nas condições apresentadas neste projeto, quando comparado com MAXVER.

Logo, levando em consideração os resultados obtidos, vê-se uma contradição, uma vez que os resultados apresentados nas Figura 6 e 7 sugerem como mais eficaz o classificador por MAXVER, enquanto o Índice Kappa sugere que seja por RF. Contudo, compreendendo o cenário ideal como um classificador que apresente Índice Kappa entre o intervalo de 0,80 – 1,00, que corresponde a uma concordância quase perfeita, subentende-se que os classificadores de MAXVER e RF podem não ser as melhores opções para a finalidade de identificar

indivíduos de *Pinus* de forma isolada. Além disso, a ineficácia dos classificadores também está associada à qualidade dos equipamentos utilizados para obtenção das fotografias aéreas e à qualidade das amostras utilizadas para treinamento dos classificadores, amostras estas que estão diretamente relacionadas aos resultados da classificação.

Por essas razões, sugere-se a busca por outros classificadores que possam ter melhores desempenhos nas condições dos materiais apresentados neste projeto.

4. CONCLUSÕES

Como proposto nos objetivos, a ferramenta de comparação utilizada neste projeto foi a matriz de confusão, mais especificamente o Índice Kappa. Segundo os índices obtidos com a metodologia apresentada, o classificador que se mostrou mais eficaz foi o de *Random Forest*, uma vez que seu índice foi maior e, portanto, sugere maior concordância entre a classificação real (realizada manualmente) e a classificação realizada pelo próprio classificador. Entretanto, dadas todas as variáveis envolvidas, os classificadores podem apresentar desempenhos diferentes de acordo com os equipamentos utilizados e a quantidade e qualidade das amostras realizadas, os quais são muito importantes para a classificação.

Logo, a fim de obter melhores resultados, ressalta-se a importância de aprimorar as amostras, aumentando sua quantidade e a qualidade de sua representatividade, bem como de optar por equipamentos, como câmeras que ofereçam maior número de bandas, por exemplo, o que enriqueceria a base de dados que auxiliam o classificador na compreensão das amostras. Além disso, para a finalidade de identificar indivíduos de *Pinus sp.* em áreas de campo úmido, com os equipamentos apresentados, os classificadores de MAXVER e *Random Forest* podem não ser as melhores opções, sugerindo a busca por outros classificadores que apresentem melhores desempenhos.

REFERÊNCIAS

- BUSSAB, W.O.; MORETTIN, P.A. **Estatística Básica**. São Paulo: Saraiva, 9ª ed., 568 p., 2017.
- AMARAL, C. H. *et al.* Mapping invasive species and spectral mixture relationships with neotropical woody formations in southeastern Brazil. *ScienceDirect, ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 108, p. 80-93, Out. 2015. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0924271615001677?via%3Dihub>>. Acesso em: 12 out. 2021.
- ASNER, G. P. *et al.*, Remote sensing of native and invasive species in Hawaiian forests. **ScienceDirect**, *Remote Sensing of Environment*, vol. 112, p. 1912-1926, fev. 2007. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0034425708000199>>. Acesso em: 12 out. 2021.
- BRADLEY, B. A. Remote Detection of Invasive Plants: A Review of Spectral, Textural and Phenological Approaches. **Biol Invasions**, Amherst, MA, EUA, n. 16, p. 1411-1425, out. 2013. Disponível em: <<https://link.springer.com/article/10.1007/s10530-013-0578-9>>. Acesso em: 12 out. 2021.
- BREIMAN, L. *Random Forest, Machine Learning*, n. 45, p. 5-32, 2001.
- CÂMARA, G.; SOUZA, R.C.M.; FREITAS, U. M.; GARRIDO, J. C. P. SPRING: Integrating Remote Sensing and GIS with Object-Oriented Data Modelling. *Computers and Graphics*, v.15, n.6, p.13-22, 1996.
- CONCIANI, D. *et al.* Análise Temporal da Invasão Biológica de *Pinus sp.* em Área Úmida do Domínio Cerrado. **Revista Brasileira de Geografia Física (RBGF)**, vol. 11, n. 02, p. 521-531, 2018. Disponível em: <<https://periodicos.ufpe.br/revistas/rbgfe/article/view/234195/29379>>. Acesso em: 12 out. 2021.
- CRONK, Q.C. B.; FULLER, J.L. *Plant invaders*. London, Chapman & Hall. 1995.
- GONÇALVES, C. C. L. **Identificação automática de plantas invasoras em imagens aéreas**. 2019. 98 f. Tese (Mestrado em Engenharia de Telecomunicações e Informática) – Instituto Universitário de Lisboa, Lisboa, Portugal, 2019. Disponível em: <<https://repositorio.iscte-iul.pt/handle/10071/20313>>. Acesso em: 12 out. 2021.
- GUEDES, A. R. M.; GUIMARÃES, V. L. Sistema de Reconhecimento Baseado em *Random Forest para Caracteres de CAPTCHAS*. *Departamento de Computação, Universidade Federal de Ouro Preto, Ouro Preto, MG, 5 p., s/d. Disponível em:* <<http://www.decom.ufop.br/menotti/rp142/trab/trab1-dp2-artigo.pdf>>. Acesso em: 12 out. 2021.
- JAMES, L., EVANS, J., RALPHS, M. & CHILD, R. *Noxious range weeds*. Westview Press, Boulder. 1991.

JOSHI, C.; LEEUW, J.; DUREN, I. C. V. Remote Sensing and GIS Applications for Mapping and Spatial Modelling of Invasive Species. **XXth International Society for Photogrammetry and Remote Sensing (ISPRS) Congress: Geo-imagery Bridging Continents**, Istanbul, Turquia, p. 669-677, 12-23 Jul. 2004. Disponível em: <http://www.itc.nl/library/Papers_2004/peer_conf/joshi.pdf>. Acesso em: 12 out. 2021.

LIU, W. T. H. Aplicações de Sensoriamento Remoto. Campo Grande-MS: UNIDERP, 2006.

MATOS, D. M. S.; PIVELLO, V. R. O Impacto das Plantas Invasoras nos Recursos Naturais de Ambientes Terrestres – Alguns Casos Brasileiros. **Ciência e Cultura**, Vol. 61, n. 1, 2009. Disponível em: <http://cienciaecultura.bvs.br/scielo.php?script=sci_arttext&pid=S0009-67252009000100012>. Acesso em: 12 out. 2021.

MOONEY, H. A.; HOBBS, R. J. *Invasive species in a changing world*. Island Press, Washington. 2000.

NERY, C. V. M. *et al.* Avaliação das Técnicas de Classificação MAXVER, MAXVER – ICM e Distância Mínima Euclidiana de acordo com Índice Kappa. **Revista Brasileira de Geografia Física**, vol. 06, n. 02, p. 320-328, 2013. Disponível em: <https://www.researchgate.net/publication/281747692_Avaliacao_das_Tecnicas_de_Classificacao_MAXVER_MAXVER_-_ICM_e_Distancia_Minima_Euclidiana_de_acordo_com_Indice_Kappa_Evaluation_of_Classification_Techniques_MAXVER_MAXVER-ICM_and_Minimum_Euclidean_Dis>. Acesso em: 12 out. 2021.

REJMANEK, M. "What makes a species invasive?" In: Pysek, P., Prach, K., Rejmanek, M. & Wade, M. *Plant Invasions*. SPB Academic Publ., Amsterdam. p. 3–13, 1995.

SIQUEIRA, J. P. D. **Os conflitos institucionais da gestão florestal no Brasil - um benchmarking entre os principais produtores florestais internacionais**. 182 f. Tese (Doutorado em Engenharia Florestal) - Setor de Ciências Agrárias, Universidade Federal do Paraná, Curitiba, 2003.

SOUZA, K. R. *et al.* Classificação Automática de Imagem do Satélite Rapideye para o Mapeamento de Áreas Cafeeiras em Carmo de Minas, MG. **VII Simpósio de Pesquisa dos Cafés do Brasil**, Ago. 2011, Araxá – MG. Disponível em: <http://www.sapc.embrapa.br/arquivos/consorcio/spcb_anais/simposio7/76.pdf>. Acesso em: 12 out. 2021.

TANNUS, J. L. S. **Estudo da Vegetação dos Campos Úmidos de Cerrado: Aspectos Florísticos e Ecológicos**. 2007. 146 f. Tese (Doutorado em Biologia Geral) – Instituto de Biociências da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Rio Claro, 2007. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=171483>. Acesso em: 12 out. 2021.

VASQUES, A. G. *et al.* Uma Síntese da Contribuição do Gênero *Pinus* para o Desenvolvimento Sustentável no Sul do Brasil. **Floresta**, Curitiba, PR, vol. 37, n. 3, set/dez 2007. Disponível em: <<https://revistas.ufpr.br/floresta/article/view/9941/6834>>. Acesso em: 12 out. 2021.

WILLIAMSON, M. *Biological invasions*. London, Chapman & Hall. 1996.

XIE, Z.; ROBERTS, C.; JOHNSON, B. Object-based target search using remotely sensed data: A case study in detecting invasive exotic Australian Pine in south Florida. **ScienceDirect**, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 63, p. 647-660, mai. 2008. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0924271608000300>>. Acesso em: 12 out. 2021