

Comparativo entre os métodos de classificação MaxVer e *Random Forest* utilizando imagem *Sentinel-2B*

Uilmer Rodrigues Xavier da Cruz – Universidade Federal de Minas Gerais – (UFMG)¹

Luciel Passos de Oliveira – Universidade Federal da Bahia – (UFBA)²

Resumo: A elaboração de mapas de uso e cobertura do solo por meio de algoritmos de classificação supervisionada é uma técnica que permite o monitoramento contínuo de recursos naturais em larga escala e fornece informações valiosas. Existem vários classificadores que podem ser utilizados, cada um com suas premissas específicas. Este estudo teve como objetivo comparar o desempenho de dois classificadores distintos: a Máxima Verossimilhança (MaxVer) e a *Random Forest*. As classificações foram realizadas no sul de Minas Gerais, em uma região com áreas agrícolas e urbanas, utilizando imagem do satélite *Sentinel-2B*. A *Random Forest* obteve o melhor desempenho entre os dois classificadores, com índice Kappa de 0,77, embora tenha apresentado problemas para detectar corpos d'água menores. Portanto, trata-se de um algoritmo indicado para um mapeamento mais preciso de áreas com características diferentes.

Palavras-chave: Sensoriamento remoto; Processamento digital de imagens; Vegetação.

Comparative between MLC and Random Forest classification algorithms using Sentinel-2B image

Abstract: The elaboration of land-use and land-cover maps using classifying algorithms is a technique that allows the continuous monitoring of natural resources on large scales and provides valuable information. There are several classifiers can be used, each with its specific premises. This study aimed to compare the performance of two different classifiers: the Maximum Likelihood Classification (MLC) and the Random Forest. The classifications were carried in southern Minas Gerais, in an area with agricultural and urban sites, using image from the Sentinel-2B. The Random Forest obtained the best performance between the two classifiers, with a Kappa index of 0.77, although it presented issues to detect smaller water bodies. Therefore, it is an algorithm indicated for a more accurate mapping of areas with different characteristics.

Keywords: Remote sensing; Digital image processing; Vegetation.

¹ Doutorando em Geografia pela Universidade Federal de Minas Gerais (UFMG). Bolsista de doutorado da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) (2020). Mestre em Geografia pela Universidade do Estado do Rio de Janeiro (UERJ) (2019), Graduação em Geografia - Ênfase em Sistemas de Informações Geográficas pela Pontifícia Universidade Católica de Minas Gerais (PUC-MG) (2008). Atualmente é pesquisador do Grupo de Estudos e Pesquisas em Resíduos Sólidos - (UEMG) - - uilmer@ufmg.br

² Doutor em Geografia pelo Instituto de Geociências da Universidade Federal da Bahia (IGEO-UFBA) (2021). Mestre em Ciências Ambiental no Programa de Modelagem de Ciências da Terra e do Ambiente na Universidade Estadual de Feira de Santana (PPGM-UEFS), Licenciado e Bacharel em Geografia pela Universidade Estadual de Feira de Santana (UEFS) (2011.2 e 2014.2).

1 - INTRODUÇÃO

A transformação dos ambientes naturais causada por diversos fatores, como agricultura, uso da água e expansão urbana, vem modificando as funções ecológicas e ecossistêmicas, gerando alterações significativas no uso e na ocupação do solo (GUIMARÃES et al., 2018). Nesse sentido, as geotecnologias passam a ser ferramentas valiosas na gestão dos recursos naturais (RICHARDS; JIA, 2006), uma vez que possibilitam a coleta, o processamento, a análise e a oferta de informações com referência geográfica (ROSA, 2005). A associação do Sensoriamento Remoto (SR) aos Sistemas de Informação Geográfica (SIG) permite a realização de vários trabalhos na área das ciências ambientais, como o mapeamento do uso e da cobertura do solo, possibilitando análises em escalas multitemporais, com ampla possibilidade de extração de informações (ROSA, 2005).

Segundo Novo e Ponzoni (2001), o advento do SR está situado desde o surgimento da fotografia até os estudos espaciais. Eles comentam que “as fotografias aéreas foram o primeiro produto de sensoriamento remoto a ser utilizado, tanto é assim, que a fotogrametria e a fotointerpretação são termos muito anteriores ao termo sensoriamento remoto propriamente dito” (p. 4). São várias as definições conceituais para o SR. No entanto, os autores se referem a ele como um conjunto de técnicas destinado à obtenção de informação sobre objetos, sem que haja contato físico com estes.

Neste estudo, a plataforma de SR utilizada pertence à missão *Sentinel*, da Agência Espacial Europeia (EUROPEAN SPACE AGENCY, 2021a), que lançou seu primeiro satélite, *Sentinel-2A*, em abril de 2014, com sensores-radar para monitoramento dos recursos terrestres e oceânicos. Em junho de 2015, foram lançados à órbita da Terra os satélites gêmeos *Sentinel-2A* e *B*, com sensores multiespectrais de alta resolução espacial e temporal, para monitoramento da vegetação, dos solos e de áreas costeiras (EUROPEAN SPACE AGENCY, 2021b). As características do *Sentinel-2A* e do *Sentinel-2B* estão descritas no quadro a seguir (*Quadro 1*).

Quadro 1. Características básicas do *Sentinel-2A* e do *Sentinel-2B*.

<i>Sentinel-2A/ 2B</i>	
Sensor	MSI (<i>Multi-spectral instrument</i>)
Resolução Espacial	10m (4 bandas VNIR), 20m (6 bandas), 60m (3 bandas)
Resolução Temporal	5 dias (equador) 2-3 dias (méd. latitude)
Swath	290 km

Fonte: OLIVEIRA, 2017.

Para Barbosa e Maillard (2010), “as imagens de satélite de alta resolução, ricas em detalhes, aumentam a necessidade de técnicas de classificação mais sofisticadas” (p. 233, tradução nossa³). Com a evolução dos satélites e de suas imagens, a fotointerpretação deve ser minuciosa, uma vez que seu detalhamento será maior e demandará mais atenção tanto no processamento de imagens, quanto no método de classificação. Assim, Jensen (2011) indica que as técnicas de Processamento Digital de Imagem (PDI) busquem “[...] detectar e identificar fenômenos importantes em uma cena”, com o intuito de tornar as imagens visualmente melhores e facilitar sua análise. Para estudar uma imagem de satélite e extrair a informação do alvo, é necessário processá-la digitalmente. As técnicas de PDI tratarão os dados brutos das imagens de satélite e minimizarão os ruídos, de maneira que seja possível uma melhor interpretação dos dados.

Segundo Shiba et al. (2005), na classificação supervisionada “são utilizadas as técnicas de aprendizagem de máquina para a extração de informações a partir de conhecimento a priori do usuário” (p. 4319). Assim, são selecionados alvos, na imagem, que servirão para “treinar” o algoritmo a procurar objetos semelhantes. Os algoritmos utilizados variam de acordo com a aplicação e a estrutura de dados escolhida. Existem vários métodos para mapeamento do uso e da cobertura do solo, sendo que a principal premissa desses métodos consiste na interpretação e na classificação de imagens disponíveis. A classificação pode ser visual e manual, além de automatizada. Os classificadores automatizados são amplamente utilizados, uma vez que permitem o mapeamento contínuo de grandes extensões (LANG; BLASCHKE, 2007). De acordo com Richard e Jia (2006), as técnicas de classificação automática supervisionada por pixel podem ser uma alternativa viável para gerar mapas de uso e cobertura do solo. Essas técnicas consistem na coleta de amostras representativas de cada classe do uso, as quais correspondem a atributos do espaço espectral e funcionam como áreas de treinamento para o algoritmo classificador (FLORENZANO, 2011).

No entanto, existem vários tipos de algoritmos classificadores. Segundo Campos (2016), o classificador denominado Máxima Verossimilhança (MaxVer) é o método de classificação supervisionado mais utilizado. De acordo com Gorte et al. (1999), “o seu algoritmo calcula a probabilidade de um dado pixel pertencer a uma das classes, considerando como normal ou gaussiana a distribuição da amostra de treinamento, descrita pelo vetor de média e pela matriz de covariância do pixel” (p. 2). Outro algoritmo de classificação bastante utilizado em diversos estudos é a *Random Forest (RF)*, um algoritmo de

³ High-resolution satellite images, rich in detail, increase the necessity for more sophisticated classification techniques.

machine learning que utiliza um conjunto de árvores de decisão independentes para realizar classificações e predições com alta precisão e baixo enviesamento (BREIMAN, 2001). A RF, além de ser amplamente empregado em processos de classificação, também é útil no monitoramento, em ambientes agrícolas ou naturais, de aspectos biofísicos de plantas, tais como *status* nutricional, teor hídrico, biomassa e teor de clorofila (PRADO-OSCO et al., 2019).

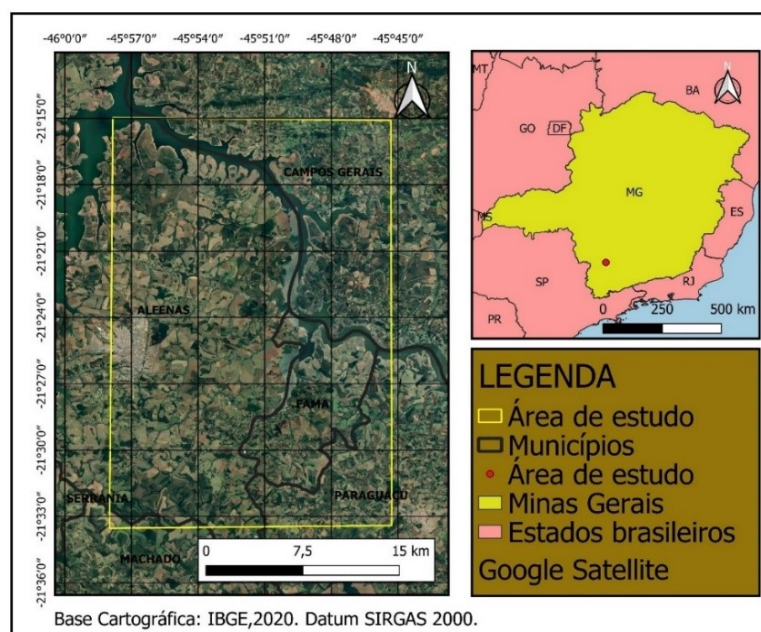
Sendo assim, com a finalidade de comparar a performance dos dois classificadores supracitados (MaxVer e RF), foram realizadas duas classificações supervisionadas de uma área com múltiplos usos da terra, localizada no sul de Minas Gerais. Os mapeamentos foram feitos utilizando uma cena, de setembro de 2020, do satélite *Sentinel-2B*, com resolução espacial de 10 m. O desempenho dos classificadores foi avaliado por meio do índice de Kappa.

2 - METODOLOGIA

2.1 - Área de estudo

A região escolhida para classificação possui uma área total de 744,77 km², que abrange parte dos municípios de Alfenas, Fama, Paraguaçu, Campos Gerais e Machado, no sul de Minas Gerais. Essa região faz parte do contexto da Bacia Hidrográfica do Entorno do Lago de Furnas (BHELF), sub-bacia do Rio Grande, e da microrregião de Alfenas (ver *Mapa 1*).

Mapa 1. Localização da área de estudo em relação ao estado de Minas Gerais.



Fonte: Elaborado pelos autores.

Trata-se de uma região onde predominam cidades pequenas e médias, cuja principal atividade econômica é a agricultura, em especial a cafeicultura (ALVES; LIDNER, 2020). De acordo com Lense et al. (2020), na região do entorno do Lago de Furnas, o espaço é predominantemente ocupado por pastagens e culturas agrícolas, com baixa ocorrência de vegetação nativa, como Mata Atlântica e Cerrado. O tipo climático que predomina na região é o Tropical de Altitude (*Cma*), caracterizado por temperaturas amenas, com invernos secos e verões úmidos (LENSE et al., 2020; MINAS GERAIS, 2013).

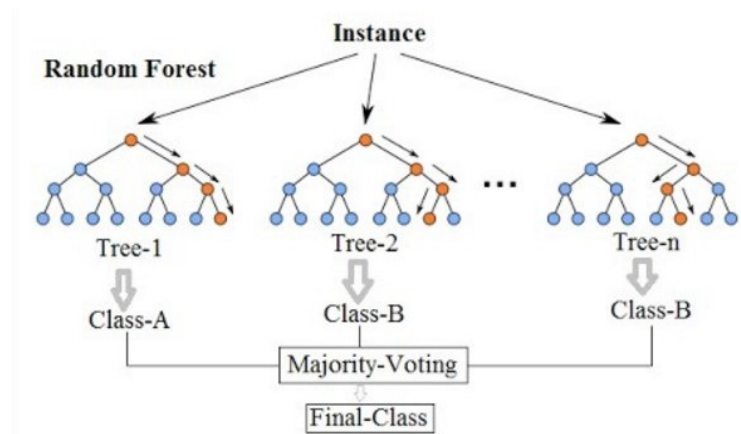
2.2 - Descrição dos classificadores utilizados

Este trabalho analisa classificações supervisionadas elaboradas utilizando os algoritmos de Máxima Verossimilhança (MaxVer) e *Random Forest* (RF). A MaxVer é um algoritmo que trabalha com pixels homogêneos e os classifica de acordo com a semelhança deles, analisando o contexto espacial da vizinhança na qual estão inseridos e calculando a probabilidade de os pixels pertencerem ou não a uma determinada classe (CECHIM JÚNIOR; JOHANN; ANTUNES, 2017).

Shiba et al. (2005) explica que “a eficácia do [algoritmo de] MaxVer depende, principalmente, de uma precisão razoável da estimativa do vetor médio e da matriz de covariância de toda classe espectral” (p. 4320). Dessa forma, a acurácia dele estará relacionada à quantidade de pixels nas amostras de treinamento que abastecem a matriz de covariância (SHIBA et al., 2005). A MaxVer, que é um classificador paramétrico, é o método mais utilizado em classificações de sensoriamento remoto, sendo empregada em cerca de 32% de todos os artigos (YU et al., 2014).

A *Random Forest*, por sua vez, é um algoritmo de *machine learning* que utiliza um conjunto de árvores de decisão para realizar classificações e estimativas por meio de regressão (MAXWELL; WARNER; FANG, 2018). O algoritmo divide o conjunto de dados em dois subconjuntos, um de treinamento e outro de teste, empregando o método *bagging*. O subconjunto de treinamento, com cerca de dois terços dos dados, é usado, justamente, para treinar o modelo e, em seguida, realizar as classificações no subconjunto de teste. A classificação final é feita considerando a classe mais votada dentre as árvores de decisão da “floresta” (ver *Figura 1*).

Figura 1. Ilustração simplificada do processo de classificação *Random Forest*.



Fonte: KOHERSEN, 2017.

2.3 - Definição das classes de uso e ocupação

A imagem *Sentinel-2B* da área foi obtida, de forma gratuita, pela plataforma *Earth Explorer*, do Serviço Geológico dos Estados Unidos (USGS), através do canal de distribuição *on-line* <earthexplorer.usgs.gov>, por onde foi selecionada uma cena sem nuvens datada em 8 de setembro de 2020, final do período de estiagem da região. Para a determinação das classes que seriam mapeadas, foram considerados os dados da mais recente pesquisa de Produção Agrícola Municipal (PAM), do Instituto Brasileiro de Geografia e Estatística (IBGE), referente ao ano de 2019. De acordo com a PAM de 2019, a microrregião de Alfenas possuía mais de 85 mil ha de lavouras temporárias e mais de 72 mil ha de lavouras perenes. Em relação ao total da área ocupada por culturas temporárias, destacam-se a produção de milho (40,6%), soja (26,7%), feijão (16,5%) e cana (11%). Em menor proporção, também são produzidos na região batata inglesa, batata-doce, trigo, sorgo, tomate, mandioca, girassol, arroz, aveia etc. (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2020).

Já no que diz respeito às lavouras permanentes, destaca-se a produção de café-arábica, com 85% do total de área plantada de culturas perenes, embora a região também produza laranja, banana, limão, tangerina, abacate, entre outras frutas, cultivadas, principalmente, em pequenas áreas de agricultura familiar. Além disso, de acordo com Lense et al. (2020), a região também possui uma alta ocorrência de pastagens, superior a 35% do total da Bacia Hidrográfica do Entorno do Lago de Furnas (BHELFF). Dessa forma, no que diz respeito às culturas agrícolas, foram definidas quatro classes principais: i) lavouras temporárias, representada, principalmente, pelas áreas de produção de milho, soja e cana; ii) café, que ocupa a maior parte das lavouras perenes, sendo mais reconhecível por

imagens na resolução de 10 m, como é o caso deste estudo; iii) eucalipto, que é facilmente reconhecível pelo padrão espacial de distribuição dos talhões; e iv) pastagens.

Ademais, também foram mapeadas as seguintes classes: corpos hídricos, áreas urbanas, solo exposto, cicatrizes de queimadas e fragmentos de mata, totalizando nove classes de uso e ocupação. Embora entenda-se que as áreas queimadas sejam comuns em áreas de cultivo de lavouras temporárias, como a cana-de-açúcar, estas foram consideradas uma classe à parte, tanto pela ausência de dados precisos em relação às lavouras, quanto pela ocorrência de queimadas durante o período de estiagem.

2.4 - Aquisição, processamento e fotointerpretação da imagem *Sentinel-2B*

Inicialmente, por meio da ferramenta Miscelânea do software *QGIS (Qgis Development Team)*, foi feito o empilhamento de bandas e, em seguida, gerado uma composição falsa-cor utilizando as bandas 8 (infravermelho próximo), 4 (vermelho) e 3 (verde). A classificação foi realizada extraindo entre 5 e 10 polígonos amostrais representativos das classes de interesse, que foram usados posteriormente para treinar os algoritmos. As áreas urbanas, por serem compostas por uma grande variedade de objetos, tais como asfalto, telhas, água, vegetação, solo, metal etc., possuem um padrão espectral complexo, similar ao de materiais minerais (JENSEN, 2011).

As lavouras agrícolas, em especial as temporárias, são afetadas pelos ciclos de plantio e colheita, que variam de cultura para cultura. Por exemplo, o plantio de soja ocorre entre outubro e dezembro, e suas áreas são, após a colheita, destinadas à produção de feijão, milho etc. Dessa forma, a presença da vegetação não foi o elemento mais importante na amostragem dessas áreas, e sim o padrão espacial dos talhões, uma vez que vários destes se encontravam em fase pós-colheita ou pré-plantio. Assim, durante a classificação inicial, foram geradas duas classes distintas para representar as lavouras temporárias, que, por fim, foram reclassificadas como apenas uma classe homogênea.

As pastagens também foram classificadas considerando suas particularidades, uma vez que se diferenciam pela ausência de formatos geométricos dos talhões e pela presença de árvores esparsas e zonas de erosão pelo pisoteio do gado. A classificação das áreas de eucalipto fica facilitada pela organização espacial em talhões geométricos ou quase geométricos, pela alta densidade de plantio e pela diferente textura aparente em relação às áreas de mata. Uma vez coletadas as amostras em forma de polígonos, o processo de classificação se deu pelos seguintes passos: i) cálculo das estatísticas das imagens; ii) treinamento e geração dos modelos; iii) classificação da imagem; e iv) regularização do

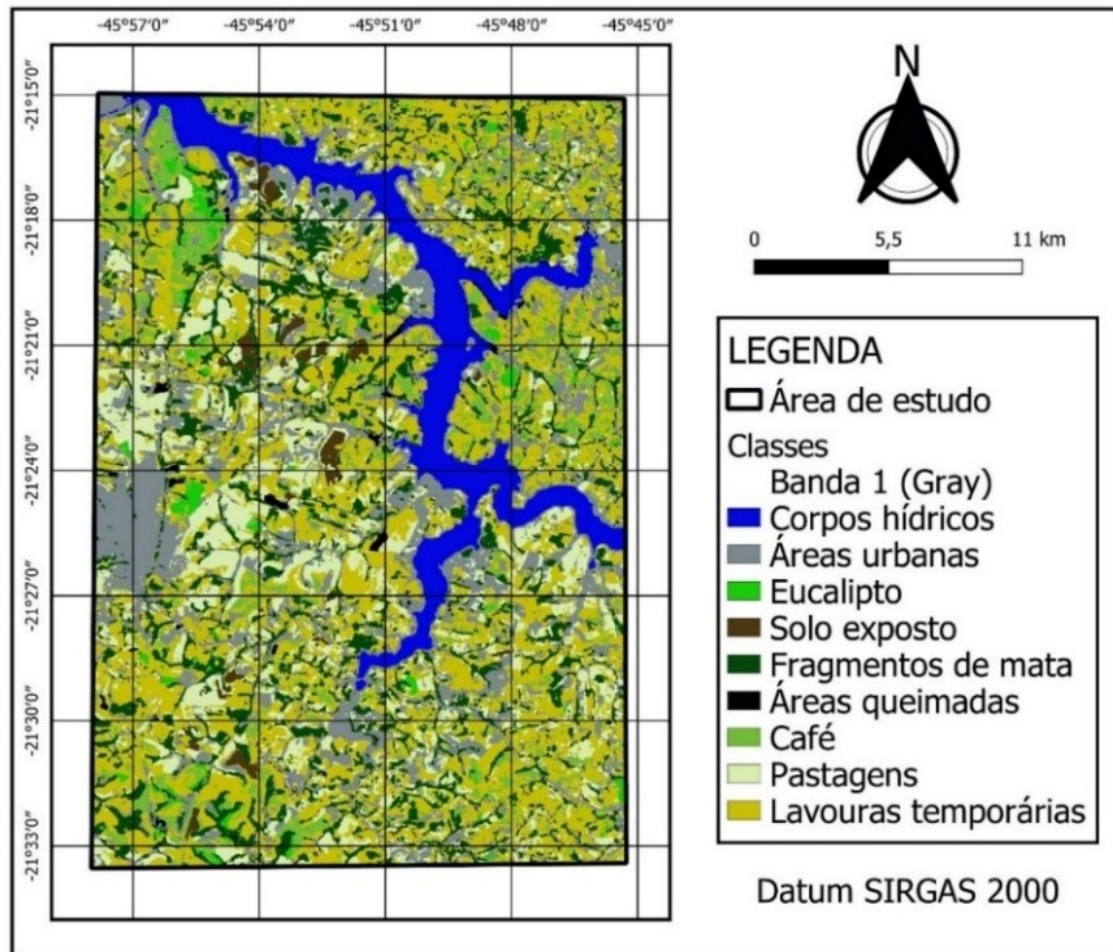
mapa de classificação. Todos os processos foram realizados utilizando o *plug-in Orpheu Toolbox (OTB)* no *QGIS*, com os classificadores *rf* para o modelo *Random Forest* e *bayes*, para a *MaxVer*.

O provedor *GRASS* do *QGIS* foi utilizado para realizar o cálculo da área de ocupação de cada classe, utilizando a função *r.report*. Por fim, como modo de avaliar e comparar o desempenho dos dois métodos de classificação, foi calculado, também no *QGIS*, o coeficiente de Kappa dos dois mapas resultantes. O coeficiente de Kappa (K), proposto em 1960 por Jacob Cohen, mensura o grau de concordância entre “proporções derivadas de amostras dependentes” (SILVA; PAES, 2012), indicando a qualidade do modelo, ou seja, do mapa resultado.

3 - RESULTADOS E DISCUSSÃO

A partir dos perecimentos envolvendo o Processamento Digital de Imagem (PDI) e as análises dos dados e das informações secundárias, foi possível elaborar um conjunto de mapas de uso e cobertura das terras com os dados dos sensores remotos utilizados no estudo. Os mapas foram elaborados empregando algoritmos de classificação distintos (*machine learning Random Forest* e Máxima Verossimilhança). O mapa a seguir (*Mapa 2*) espacializa o resultado da classificação pelo método de Máxima Verossimilhança, empregando nove classes de uso e cobertura das terras na área investigada por esta pesquisa. Já a tabela subsequente (*Tabela 1*) registra a proporção de área ocupada, na área de estudo, por cada uma das nove classes de interesse.

Mapa 2. Mapa de uso e cobertura da terra gerado a partir da classificação por Máxima Verossimilhança.



Fonte: Elaborado pelos autores.

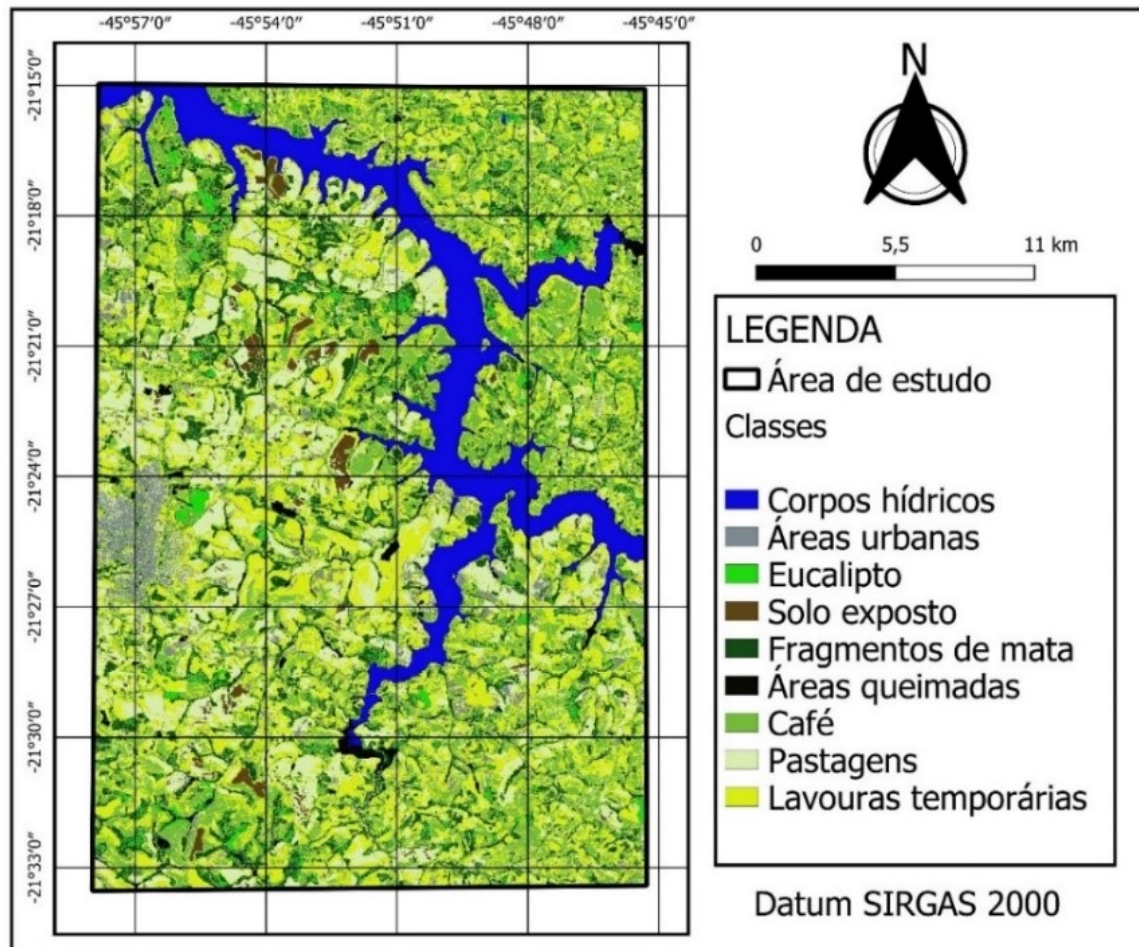
Tabela 1. Proporção de ocupação de cada classe, na área de estudo, com a classificação MaxVer.

Classe	Nº pixels	% Ocupação	Área (km ²)
Corpos hídricos	565596	7,61	56,68
Área urbana	1575516	21,19	157,82
Eucalipto	110566	1,49	11,10
Solo exposto	91049	1,22	9,09
Fragmentos de mata/ciliar	974641	13,11	97,64
Áreas queimadas	44361	0,60	4,47
Café	492482	6,62	49,31
Pastagens	1029456	13,84	103,05
Lavouras temporárias	2551930	34,32	255,61
Total	7435597	100	744,77

Fonte: Elaborado pelos autores.

O *Mapa 3* demonstra o resultado espacializado da classificação utilizando o algoritmo de *machine learning Random Forest*. Nesse mapa, é possível notar algumas pequenas diferenças visuais, quando comparado ao mapa gerado pelo algoritmo anterior (Máxima Verossimilhança). Paralelamente, a *Tabela 2* apresenta a proporção de área ocupada, na região de estudo, por cada uma das nove classes de interesse.

Mapa 3. Mapa de uso e cobertura da terra a partir da classificação empregando o algoritmo *Random Forest*.



Fonte: Elaborado pelos autores.

Tabela 2. Porcentagem de ocupação de cada classe, na área de estudo, com a classificação *Random Forest*.

Classe	Nº pixels	% Ocupação	Área (km ²)
Corpos hídricos	685635	9,22	68,67
Área urbana	263444	3,54	26,36
Eucalipto	137369	1,85	13,78
Solo exposto	91955	1,24	9,24

Fragmentos de mata/ciliar	987756	13,28	98,91
Áreas queimadas	150761	2,03	15,12
Café	1945411	26,16	194,83
Pastagens	1126927	15,16	112,91
Lavouras temporárias	2046339	27,52	204,96
Total	7435597	100,00	744,77

Fonte: Elaborado pelos autores.

Os corpos hídricos possuem um padrão singular de resposta espectral, sendo que a água limpa reflete, principalmente, a cor azul e absorve quase a totalidade dos demais comprimentos de onda. Contudo, reservatórios hídricos naturais não são compostos apenas por água, mas também por sedimentos oriundos da erosão dos solos, escoados superficialmente pela drenagem pluvial, e ainda por matéria orgânica e vegetação e microrganismos aquáticos. A presença desses materiais na água gera diferentes colorações na composição falsa-cor, principalmente devido à atividade fotossintética, tornando possível identificar áreas de emissão de efluentes, por exemplo (JENSEN, 2011).

Os dois classificadores estimaram a presença de corpos hídricos com uma diferença de 21% entre eles (MaxVer=7,61% e RF=9,22%). Lense et al. (2020), utilizando dados do Projeto MapBiomass, estimaram que, na região da Bacia Hidrográfica do Entorno do Lago de Furnas (BHELFB), cerca de 8% é ocupada por corpos hídricos. Ambos os algoritmos se equivocaram em classificar áreas aquáticas de menor dimensão e trechos de borda do Lago de Furnas, sendo que a MaxVer confundiu principalmente essas áreas com zonas urbanas.

O Lago de Furnas tem como característica marcante a ausência de mata ciliar em suas margens (LENSE et al., 2020). Sendo assim, durante os períodos de redução do volume do reservatório (secas), as áreas de solos expostos ficam mais evidentes nas bordas do espelho d'água. A mata e/ou vegetação de margem, quando presentes, além de serem extremamente importantes para a proteção do lago contra a deposição de sedimentos erodidos, poderiam atuar na classificação como uma borda limítrofe entre o lago e as áreas agrícolas e de solo exposto, o que poderia contribuir para um melhor resultado. A RF, por sua vez, embora tenha confundido alguns trechos de água com áreas queimadas — o que pode ser justificado pela aparência escura na composição usada para o estudo —, conseguiu identificar corpos hídricos menores, incluindo os que se encontram nas margens recortadas da represa, com maior eficiência.

Já na área urbana identificada, a diferença entre os classificadores foi sensível, uma vez que a MaxVer estimou quase 6 vezes mais áreas urbanas. Como foi mencionado, o algoritmo de MaxVer gerou uma alta ocorrência de áreas classificadas como urbanas que pertenciam à classe de corpos hídricos. Áreas urbanas possuem um padrão complexo de resposta espectral, uma vez que sua composição é igualmente complexa. Dessa forma, sugere-se que, em futuras aplicações, uma máscara seja utilizada para remover as áreas urbanas das classificações, ou que o número de amostras seja superior.

No que diz respeito à classificação das áreas de eucalipto e solo exposto, a avaliação dos dados mostra que não houve diferença significativa entre os dois métodos, seja estatística, seja espacial. Ambas as classes possuem características morfológicas e espectrais que facilitam sua identificação: o solo exposto aumenta sua refletância conforme aumenta o comprimento de onda (JENSEN, 2011), adquirindo coloração quase branca em composições com a presença do infravermelho próximo, especialmente em períodos de estiagem, como foi o caso deste estudo; já o eucalipto apresenta, como já foi mencionado, uma distribuição quase geométrica dos talhões, além de alta densidade de plantio, o que facilmente o distingue de outras culturas perenes, bem como das áreas de mata.

A performance dos classificadores na identificação de áreas de fragmentos de mata foi quase idêntica, podendo ser visualizada pelo encaixe da mata ciliar em áreas de córregos, possibilitando uma clara identificação da rede de drenagem. Contudo, a MaxVer classificou como mata uma parte das áreas de cafeicultura, o que pode se justificar pela arquitetura do dossel dos cafeeiros, com maior quantidade de biomassa foliar, sobretudo em regiões de plantio adensado. A exploração de novas composições coloridas, envolvendo, por exemplo, o infravermelho médio, poderia ajudar a melhorar a distinção entre as classes. Isso se justifica porque essa região do espectro eletromagnético, em estudos de vegetação, apresenta dois canais de absorção de radiação, sendo maior quanto maior for o teor de umidade presente nas plantas (até um platô) (PONZONI; SHIMABUKURO; KUPLICH, 2015).

Como as áreas de mata são essencialmente de mata ciliar, configurando áreas de preservação permanente (APP), é razoável supor que sejam áreas menos suscetíveis ao estresse hídrico da estiagem do que os cafeeiros. Assim, podemos concluir que, para as classes eucalipto, solo exposto e fragmentos de mata, o desempenho foi excelente com ambos os métodos propostos. Tendo em vista que as taxas de semelhança são muito próximas, podemos visualizar o traçado de forma muito clara em ambos os mapas.

Em relação às áreas de queimadas, geramos os seguintes dados: 0,60% para a MaxVer e 2,03%, para a RF. Como abordado, a RF superestimou as áreas de queimada, com confusões dentro de áreas pertencentes aos corpos hídricos. Assim, especificamente para essa classe, a MaxVer obteve o desempenho que mais se aproximou da realidade. Sobre o café, a diferença entre os métodos também foi importante, com a RF estimando quase 4 vezes mais área de cafeicultura do que a MaxVer. De acordo com a PAM, a microrregião de Alfenas possui cerca de 14% do seu território ocupada por cafeicultura (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2020).

No que se refere às áreas de pastagem, houve pouca diferenciação entre os resultados finais obtidos, sendo 13,84% para a MaxVer e 15,16% para a RF. Em relação às áreas destinadas às culturas temporárias (soja, milho, cana etc.), as classificações foram uniformes, entre 20,63% para a MaxVer e 19,92% para a RF, no que diz respeito ao total da área classificada. Segundo dados da PAM, a microrregião de Alfenas possui cerca de 17% de sua área ocupada por lavouras temporárias (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2020).

4 - COMPARAÇÃO DA PERFORMANCE DOS CLASSIFICADORES POR MEIO DO COEFICIENTE DE KAPPA

Na *Tabela 3* são apresentados os valores sugeridos por Landis e Koch (1977), adaptado por Nadas (2017), para avaliar a qualidade do mapa temático. Foram coletados pontos amostrais de acuracidade ao longo da área mapeada. Neste estudo, os índices foram calculados através do *software QGIS*. Assim, a qualidade do mapa resultante da classificação *Random Forest* obteve qualidade muito boa, com o valor de 0,77 de concordância. Já o mapa da classificação do método de Máxima Verossimilhança possibilitou uma representação com concordância inferior ao alcançado pelo algoritmo anterior (*Random Forest*). Sendo assim, a Máxima Verossimilhança trouxe o valor correspondente a 0,42 no índice de Kappa. Ao comparar os dois métodos de classificação supervisionada, percebe-se que o método *Random Forest* apresenta um valor mais satisfatório que a MaxVer, justificando a preferência por ele.

Tabela 3. Índice Kappa de acordo com o método de classificação supervisionada.

Método	Índice Kappa
Máxima Verossimilhança	0,42
Random Forest	0,77

Fonte: Elaborado pelos autores.

O método de classificação pela Máxima Verossimilhança (MaxVer), por presumir uma distribuição normal dos dados, apresenta melhores resultados com dados unimodais, demonstrando limitações com dados em que há maior variabilidade intraclasses (LIU; SHI; ZHANG, 2011). Essas limitações têm provocado o crescimento do uso de algoritmos poderosos de *machine learning*, tal como a *RF*. Contudo, mesmo que esses algoritmos apresentem maior acurácia em processos de classificação, a MaxVer ainda é o método mais utilizado (MAXWELL; WARNER; FANG, 2018; YU et al., 2014).

A principal vantagem do processo de classificação supervisionada é a de que o analista tem maior controle ao determinar classes e identificar, por meio das análises, possíveis ruídos nas amostras de treinamento. Por outro lado, o analista define uma estrutura de classificação, *a priori*, à imagem, a partir da seleção prévia das classes de informação, ou seja, as classes determinadas podem não coincidir com a realidade no campo (BARBOSA et al., 2018).

Patsalidis, Agapiou e Hadjimitsis (2019) compararam o desempenho dos classificadores *RF* e MaxVer, usando imagens *Sentinel-2* e *Landsat-8*, em um ambiente semiárido do Mediterrâneo. Assim como neste estudo, eles obtiveram maior acurácia com a *RF*, que alcançou um coeficiente de Kappa correspondente a 0,85 de acerto, enquanto a MaxVer ficou com valores entre 0,71 e 0,79 de acerto, considerando uma escala de 0,01 (baixa precisão) até 1,0 (alta precisão) (LANDIS; KOCH, 1977). Com o modelo *RF*, os autores indicam que as classes com menor acurácia foram às áreas urbanas e a vegetação de baixa intensidade, enquanto a classificação de corpos hídricos naturais foi 100% acurada em todos os bancos de dados utilizados.

Silva et al. (2020) compararam a performance do classificador MaxVer com um poderoso algoritmo de *machine learning* — a Rede Neural Artificial (RNA) —, para mapear o uso e a ocupação do solo em áreas de cerrado dos estados de Goiás e Minas Gerais, no Brasil, também com o uso de imagens *Sentinel-2*. O método RNA foi ligeiramente mais eficiente que a MaxVer, obtendo um coeficiente de Kappa de 0,77. O uso do Índice de Vegetação pela Diferença Normalizada (NDVI) gerou melhores modelos. No que diz respeito às classes, o mapeamento de pastagens e formações de savana gerou maiores erros de classificação, enquanto corpos hídricos e formações florestais foram mapeadas com maior acurácia.

Wasniewski et al. (2020) utilizaram a *RF* e imagens *Sentinel-2* para classificar cinco tipos de florestas tropicais no Gabão. Eles alcançaram altos índices de acurácia geral e altos coeficientes de Kappa (entre 0,83 e 0,89), em especial quando inseriram o modelo digital de

elevação (MDE) no modelo de predição. O MDE foi diferencial nos modelos de Wasniewski et al. (2020) porque as ocorrências de alguns tipos de florestas são altamente dependentes da elevação. Assim, para futuros estudos, adicionar um modelo de elevação pode colaborar para reduzir erros de classificação, como os obtidos neste estudo. Como exemplo, a elevação pode ajudar a minimizar a confusão entre as áreas urbanas e as bordas não vegetadas da represa, bem como melhorar a classificação das áreas de cafeeiros, tradicionalmente plantadas em áreas de maior elevação. Os dados *Shuttle Radar Topography Mission (SRTM)* estão disponíveis gratuitamente por meio da plataforma *Earth Explorer*.

O mapeamento de áreas de vegetação esparsa e pouco densa pode gerar erros de classificação (SILVA et al., 2020; PATSALIDIS; AGAPIOU; HADJIMITSIS, 2019), como foi o caso também deste estudo. Portanto, ferramentas como os índices de vegetação (NDVI e outros) e os modelos digitais de elevação podem ser combinadas, juntamente com os dados espectrais das bandas, para gerar melhores predições e, conseqüentemente, melhores mapeamentos. Entretanto, apenas o tratamento das imagens através das ferramentas do Processamento Digital de Imagem (PDI) não é suficiente para uma boa classificação. Baker et al. (2006) afirmam que o inventário desses ambientes depende principalmente de trabalho de campo aéreo, interpretação fotográfica e PDI. Esse último geralmente demanda menos tempo para informações, aquisição e produção de mapas e, muitas vezes, é considerado a técnica mais eficiente (MELACK, 2004). Uma alternativa que pode colaborar para aprimorar a precisão das classificações é utilizar, em combinação com as imagens de satélite, veículos aéreos não-tripulados (VANT), atualmente designados como Drones. Os Drones são ferramentas poderosas para mapeamento, uma vez que, em geral, são de baixo custo de aquisição e de fácil transporte e operação. Ademais, produzem aerofotos de altíssima resolução espacial (casa dos cm) e precisão geográfica, que são facilmente combinadas com os Sistemas de Informação Geográfica (SIG). Os Drones são ideais para áreas menores (MANFREDA et al., 2018).

Parreiras et al. (2020) utilizaram índices de vegetação oriundos de imagens de VANT, de altíssima resolução, para classificar áreas de cultivo de café, com presença de plantas invasoras, por meio do método da Máxima Verossimilhança. Obtiveram um coeficiente de Kappa de 0,92, o que, segundo eles, denota um nível de concordância quase perfeito entre modelo e dados de campo. Infelizmente, devido à pandemia de COVID-19, neste presente estudo não foi possível realizar trabalhos de campo e fazer o inventário de aerofotos ortorretificadas para validar as amostras utilizadas no treinamento das classificações ou aprimorar os resultados. Porém, no intuito de calibrar, validar os dados

requisitamos o auxílio do trabalho de campo virtual proporcionado pelo sensoriamento remoto através da web plataforma *Google Maps*, utilizando em alguns pontos a ferramenta *Google Street View* - alta precisão de 1 a 20 metros, na qual, conferimos em torno de 40 pontos estratégicos das amostras coletadas de cada classe.

Como sugestão final, com o objetivo de reduzir os erros, pode-se realizar um processo maior de coleta de amostragens no momento da classificação da imagem, bem como usufruir das séries temporais disponíveis, tanto para *Sentinel-2* quanto para *Landsat-8*, e realizar as classificações em diferentes períodos, para avaliar o efeito da sazonalidade. Além disso, é possível agregar dados de elevação e índices de vegetação. É importante salientar que a coleta de amostras das classes é uma etapa de grande relevância na elaboração da classificação supervisionada, uma vez que esta ampara o treinamento dos algoritmos classificadores, que, por sua vez, são responsáveis por classificar toda a área de estudo por meio de sua técnica específica. Nesse sentido, é importante que as classes sejam coerentes, a fim de possibilitar um melhor entendimento da resposta espectral de cada classe e elevar a precisão do mapeamento.

4 - CONSIDERAÇÕES FINAIS

Foi possível concluir que a aplicação dos classificadores é bem satisfatória no que tange à elaboração do mapeamento de uso e cobertura do solo, utilizando a coleta de classes pré-estabelecidas. Entretanto, ao comparar os resultados dos dois métodos, observou-se algumas discrepâncias no que diz respeito às classes. Para a escolha do algoritmo classificador, recomenda-se que verifique qual representa melhor as classes encontradas na área de estudo. No estudo em questão, o classificador *Random Forest* foi o que apresentou melhores resultados.

5 – REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, F. D.; LINDNER, M. Agronegócio do café no sul de Minas Gerais: territorialização, mundialização e contradições. *OKARA: Geografia em Debate*, João Pessoa, v. 14, n. 2, p. 433–451, 2020.

BAKER, C. et al. Mapping wetlands and riparian areas using Landsat ETM+ imagery and decision-tree-based models. *Wetlands*, Amsterdam, v. 26, n. 2, p. 4654–4674.

BARBOSA, I. S.; MAILLARD, P. Mapping a wetland complex in the Brazilian savannah using an Ikonos image: assessing the potential of a new region-based classifier. *Canadian Journal of Remote Sensing*, Milton Park, v. 36, n. 2, p. S231–S242, 2010.

BELGIU, M.; DRAGUT, L. Random Forest in remote sensing: a review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, Amsterdam, v.114, p. 24–31, 2016.

BREIMAN, L. Random Forests. *Machine Learning*, Amsterdam, v. 45, p. 5–32, 2001.

CAMPOS, W. W. et al. Avaliação de classificadores de imagem de satélite a partir do uso de uma técnica de votação. *Revista Brasileira de Cartografia, Uberlândia*, v. 68, n. 8, p. 1653–1664, 2016.

CECHIM JÚNIOR, C.; JOHANN, J. A.; ANTUNES, J. F. G. Mapping of sugarcane crop area in the Paraná State using Landsat/TM/OLI and IRS/LISS-3 images. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 21, n. 6, p. 427–432, 2017. Disponível em: <<https://www.scielo.br/j/rbeaa/a/YfCJzrj4GcqhWDCWBpQ9HGF>>. Acesso em: 12 mar. 2021.

EUROPEAN SPACE AGENCY. Sentinel overview. *Copernicus*, 2021a. Disponível em: <<https://sentinels.copernicus.eu/web/sentinel/home>>. Acesso em: 12 mar. 2021.

EUROPEAN SPACE AGENCY. Sentinel-2. *Copernicus*, 2021b. Disponível em: <<https://sentinel.esa.int/web/sentinel/missions/sentinel-2>>. Acesso em: 12 mar. 2021.

FLORENZANO, T. G. *Iniciação em Sensoriamento Remoto*. São Paulo: Oficina de Textos, 2011.

GORTE, B. G. H. et al. Supervised image classification. In: STEIN, A.; VAN DER MEER, F.; GORTE, B. G. H. (Orgs.). *Spatial statistics for remote sensing*. Dordrecht: Kluwer Academic Publishers, 1999, p. 1–14.

GUIMARÃES, L. F. et al. O uso de infraestruturas verde e azul na revitalização urbana e na melhoria do manejo das águas pluviais: o caso da sub-bacia do Rio Comprido. *Paisagem e Ambiente*, São Paulo, n. 42, p. 75–95, 2018. Disponível em: <<https://www.revistas.usp.br/paam/article/view/150703>>. Acesso em: 12 mar. 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Produção Agrícola Municipal – PAM: 2019. *IBGE*, 2020. Disponível em: <<https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9117-producao-agricola-municipal-culturas-temporarias-e-permanentes.html>>. Acesso em: 9 abr. 2021.

INSTITUTO NACIONAL DE PESQUISA ESPACIAL. Landsat. *INPE*, 2021. Disponível em: <<http://www.dgi.inpe.br/documentacao/satelites/landsat>>. Acesso em: 12 mar. 2021.

JENSEN, J. R. *Sensoriamento Remoto do Ambiente: uma perspectiva em recursos terrestres*. São José dos Campos: Parêntese, 2011.

KOHERSEN, W. Random Forest Simple Explanation. *Medium*, 2017. Disponível em: <<https://williamkoehrsen.medium.com/random-forest-simple-explanation-377895a60d2d>>. Acesso em: 9 abr. 2021.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics*, v. 33, n. 1, p. 159–174, 1977.

LANG, S.; BLASCHKE, T. *Análise da Paisagem com SIG*. São Paulo: Oficina de Textos, 2007.

LENSE, G. H. E. et al. Simulating the effect of permanent preservation areas on soil erosion rates. *Cerne*, Lavras, v. 26, n. 2, p. 193–201, 2020.

LIU, K.; SHI, W.; ZHANG, H. A fuzzy topology-based maximum likelihood classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, Amsterdam, v. 4, n. 7, p. 103–114, 2010.

MANFREDA, S. et al. On the use of Unmanned aerial systems for environmental monitoring. *Remote Sensing*, Basel, v. 10, n. 641, 2018.

MAWXELL, A. E.; WARNER, T. A.; FANG, F. Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing*, v. 39, n. 9, p. 2784–2817, 2018.

MELACK, J. M. Remote sensing of tropical wetlands. In: USTIN, S. (Org.). *Manual of Remote Sensing*. New York: John Wiley, 2004, p. 329–343.

MINAS GERAIS. *Plano Diretor de Recursos Hídricos da Bacia Hidrográfica do Entorno do Lago de Furnas*. Itajubá: Fundação de Pesquisa e Assessoramento à Indústria, 2013. Disponível em: <<https://alago.org.br/imagens/image/PDRH%20GD3%20-%20Diagn%C3%B3stico.pdf>>. Acesso em: 2 mar. 2021.

NADAS, M. B. et al. Análise do desempenho do classificador automático MAXVER para uso e cobertura do solo na bacia do rio Mampituba – SC. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 18, 2017. *Anais...* São José dos Campos: INPE, 2017. Disponível em: <<http://marte2.sid.inpe.br/rep/sid.inpe.br/marte2/2017/10.27.13.48.57>>. Acesso em: 12 mar. 2021.

NOVO, E. M. L. M.; PONZONI, F. J. *Introdução ao Sensoriamento Remoto*. São José dos Campos; São Paulo: INPE, 2001.

OLIVEIRA, B. S. Satélites e sensores. INPE, 2017. Disponível em: <<http://www.dsr.inpe.br/DSR/educacao/uso-escolar-sensoriamento-remoto/material-didatico-anos-anteriores/arquivos/3-satelites-e-sensores.pdf/@@download/file/3-Satelites%20e%20Sensores.pdf>>. Acesso em: 21 maio 2021.

PARREIRAS, T. C. et al. Using unmanned aerial vehicle and machine learning algorithm to monitor leaf nitrogen in coffee. *Coffee Science*, Lavras, v. 15, 2020. Disponível em: <<http://www.coffeescience.ufla.br/index.php/Coffeescience/article/view/1736/2258>>. Acesso em: 9 abr. 2021.

PATSALIDIS, S.; AGAPIOU, A.; HADJIMITSIS, D. G. Random forest classification analysis of Sentinel-2 and Landsat-8 images over semi-arid environment in the Eastern Mediterranean. In: INTERNATIONAL CONFERENCE ON GEOGRAPHIC INFORMATION, 2019. *Proceedings...* Amsterdam: Springer, 2020. Disponível em:

<https://agile-online.org/images/conference_2019/documents/short_papers/105_Upload_your_PDF_file.pdf>. Acesso em: 9 abr. 2021.

PONZONI, F. J.; SHIMABUKURO, Y. E.; KUPPLICH, T. M. *Sensoriamento remoto da vegetação*. 2. ed. São Paulo: Oficina de Textos, 2015.

PRADO OSCO, L. et al. Predicting canopy nitrogen content in citrus-trees using Random Forest algorithm associated to spectral vegetation indices from UAV-imagery. *Remote Sensin*, Basel, v. 11, n. 2925, 2019. Disponível em: <https://mdpi-res.com/d_attachment/remotesensing/remotesensing-11-02925/article_deploy/remotesensing-11-02925.pdf>. Acesso em: 9 abr. 2021.

RICHARDS, J. A.; JIA, X. *Remote Sensing Digital Image Analysis: an introduction*. Amsterdam: Springer, 2006.

ROSA, R. Geotecnologias na geografia aplicada. *Revista do Departamento de Geografia*, São Paulo, v. 16, p. 81–90, 2005.

SHIBA, M. H. et al. Classificação de imagens de sensoriamento remoto pela aprendizagem por árvore de decisão: uma avaliação de desempenho. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12, 2005, Goiânia. *Anais...* São José dos Campos: INPE, 2005, p. 4319–4326. Disponível em: <<http://mart.sid.inpe.br/col/ltid.inpe.br/sbsr/2004/11.23.11.44/doc/4319.pdf>>. Acesso em: 12 mar. 2021.

SILVA, J. F. et al. Land use/cover mapping in Brazilian Cerrado using neural network with Sentinel-2 data. *Floresta*, Curitiba, v. 52, n. 3, p. 1430–1438, 2020.

SILVA, R. S.; PAES, A. T. Teste de concordância Kappa: por dentro da Estatística. *Educação Continuada em Saúde Einstein*, v. 10, n. 4, p. 165–166, 2012.

VALE, J. R. B. et al. Análise comparativa de métodos de classificação supervisionada aplicada ao mapeamento da cobertura da cobertura do solo no município de Medicilândia, Pará. *InterEspaço: Revista de Geografia e Interdisciplinaridade*, Grajaú, v. 4, n. 13, p. 26–44, jan./abr. 2018.

WASNIEWSKI, A. et al. Assessment of Sentinel-2 satellite images and Random Forest classifier for rainforest mapping in Gabon. *Forests*, Basel, v. 11, n. 941, p. 1–17, 2020.

YU, Le et al. Meta-Discoveries Form a Synthesis of Satellite-Based Land-Cover Mapping Research. *International Journal of Remote Sensing*, v. 35, n. 13, p. 4573–4588, 2017.

Agradecimentos

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).