

INTEGRAÇÃO DE IMAGENS ORBITAIS ÓPTICAS E SAR COM PROCESSAMENTO EM NUVEM NO MAPEAMENTO DA COBERTURA DA TERRA NO CERRADO

OPTICAL AND SAR IMAGERY INTEGRATION BASED ON CLOUD COMPUTING FOR LAND COVER MAPPING IN THE CERRADO

Angela Gabrielly Pires Silva¹

Édipo Henrique Cremon²

Giovanni de Araújo Boggione³

Fabio Corrêa Alves⁴

RESUMO

O mapeamento da cobertura da terra é de suma importância para o monitoramento ambiental e gestão territorial. Séries temporais de radar de abertura sintética (SAR) do Sentinel-1 (S-1) e o sensor óptico MSI/Sentinel-2 (S-2) fornecem condições favoráveis para o mapeamento da cobertura da terra devido às suas resoluções espectrais, espaciais e temporais. Este trabalho parte do pressuposto que a combinação entre as séries temporais do S-1 e S-2 permite maior exatidão no mapeamento da cobertura da terra no bioma Cerrado. As imagens foram classificadas utilizando o algoritmo Random Forest na plataforma de processamento em nuvem Google Earth Engine. As classificações obtidas apenas com os dados S-2 ($\kappa = 89,99$) foram melhores do que as obtidas com os dados S-1 ($\kappa = 75,78$). A eficiência da classificação aumentou ao combinar os dados de ambas as missões S-1 e S-2 ($\kappa = 93,07$). Os resultados obtidos neste trabalho sugerem que a banda do infravermelho de ondas curtas, a polarização VH dos dados SAR e os índices *cellulose absorption index* (CAI) e o Hall Cover foram as variáveis mais importantes no mapeamento da cobertura da terra do bioma Cerrado.

Palavras-Chave: Google Earth Engine; sensoriamento remoto; integração de dados orbitais; Random Forest; aprendizado de máquina.

¹Instituto Federal de Goiás – IFG (Campus Goiânia), Grupo de Estudos em Geomática - GEO.
angelagabrielly225@gmail.com ORCID: <https://orcid.org/0000-0001-7759-9806>

²Instituto Federal de Goiás – IFG (Campus Goiânia), Grupo de Estudos em Geomática - GEO.
edipo.cremon@ifg.edu.br ORCID: <https://orcid.org/0000-0003-3174-7273>

³Instituto Federal de Goiás – IFG (Campus Goiânia), Grupo de Estudos em Geomática - GEO.
giovanni.boggione@ifg.edu.br ORCID: <https://orcid.org/0000-0002-1675-6529>

⁴Instituto Nacional de Pesquisas Espaciais – INPE/Divisão de Observação da Terra e Geoinformática – DIOTG.
alves.fabioc@gmail.com ORCID: <https://orcid.org/0000-0002-2941-8393>

ABSTRACT

The land cover mapping is of great relevance for the environmental monitoring and land management. Time series from the synthetic aperture radar (SAR) of Sentinel-1 (S-1) and the MSI/Sentinel-2 (S-2) optical sensor provide promising conditions for the land cover mapping due to their spectral, spatial and temporal resolutions. Here, we explored the hypothesis that the combination of S-1 and S-2 time series allows higher accuracy in the land cover mapping in Cerrado biome. The images were classified using the Random Forest algorithm in the Google Earth Engine cloud processing platform. The classifications obtained using only the S-2 data ($kappa = 89.99$) showed higher accuracy than those with the S-1 data ($kappa = 75.78$). The classification efficiency increased by combining the S-1 and S-2 data ($kappa = 93.07$). The results found here suggest that the shortwave infrared band, the VH polarization from SAR data, and the Cellulose absorption index (CAI) and Hall Cover index were the most significant variables in the land cover mapping of the Cerrado biome.

Keywords: Google Earth Engine; remote sensing; orbital data integration; Random Forest; machine learning.

INTRODUÇÃO

Uma das aplicações significativas para o levantamento de informações territoriais é o uso de imagens orbitais para o mapeamento da cobertura da terra, uma vez que, imagens tomadas destas plataformas conseguem mapear vastas áreas com qualidade e periodicidade a um baixo custo operacional. A análise de séries temporais baseada apenas em dados ópticos pode inviabilizar o monitoramento contínuo da cobertura da terra, principalmente considerando períodos chuvosos com a presença de nuvens e ou queimadas nas imagens originais. Nessas condições, o mapeamento é comprometido pela impossibilidade de interpretação do produto devido a confusão entre alvos. Uma alternativa são os sensores ativos, como os a bordo de radares imageadores, que têm como principal vantagem comparada aos sensores ópticos, a sua independência em relação à radiação solar. Além disso, os sensores ativos na faixa do espectro eletromagnético das micro-ondas possuem interação com a geometria e constante dielétrica dos alvos.

O mapeamento, a caracterização e a análise da dinâmica da cobertura da terra no bioma Cerrado, por meio de um monitoramento preciso e sistemático, utilizando o sensoriamento remoto são fatores cruciais para lidar com o desafio de ordenar a ocupação de um território de mais de dois

milhões de km², equivalentes ~ 23,92% do território nacional (IBGE, 2004). No entanto, a maioria dos estudos com esse propósito fez uso apenas de imagens ópticas (e.g. BRASIL, 2015; SOUZA *et al.*, 2020).

O uso de imagens combinadas (dados ópticos e radar de abertura sintética - SAR) para mapeamentos da cobertura da terra pode apresentar avanços na discriminação de alvos da superfície terrestre (PEREIRA *et al.*, 2016; SYMEONAKIS *et al.*, 2018; ZHANG *et al.*, 2019; LOPES *et al.*, 2020). Especificamente para o bioma Cerrado é comum a confusão entre classes de coberturas devido à similaridade espectral entre algumas classes temáticas, como a vegetação natural de formação gramíneo-lenhosas e as pastagens (SANO; JESUS; BEZERRA, 2001). A avaliação da integração de dados SAR e ópticos, usando os satélites das missões de radar do Sentinel-1 (S-1) e óptico com o sensor MSI/Sentinel-2 (S-2) com o objetivo de melhorar a detecção de classes de cobertura da terra por meio de classificadores é ainda um campo pouco pesquisado para o bioma Cerrado. Além disso, esse tipo de análise pode fornecer novas metodologias com potencial para serem reproduzidas em projetos de mapeamentos deste e de outros biomas do território nacional, como o que vem sendo executado pelo projeto TerraClass Cerrado (BRASIL, 2015) e MapBiomias (SOUZA *et al.*, 2020).

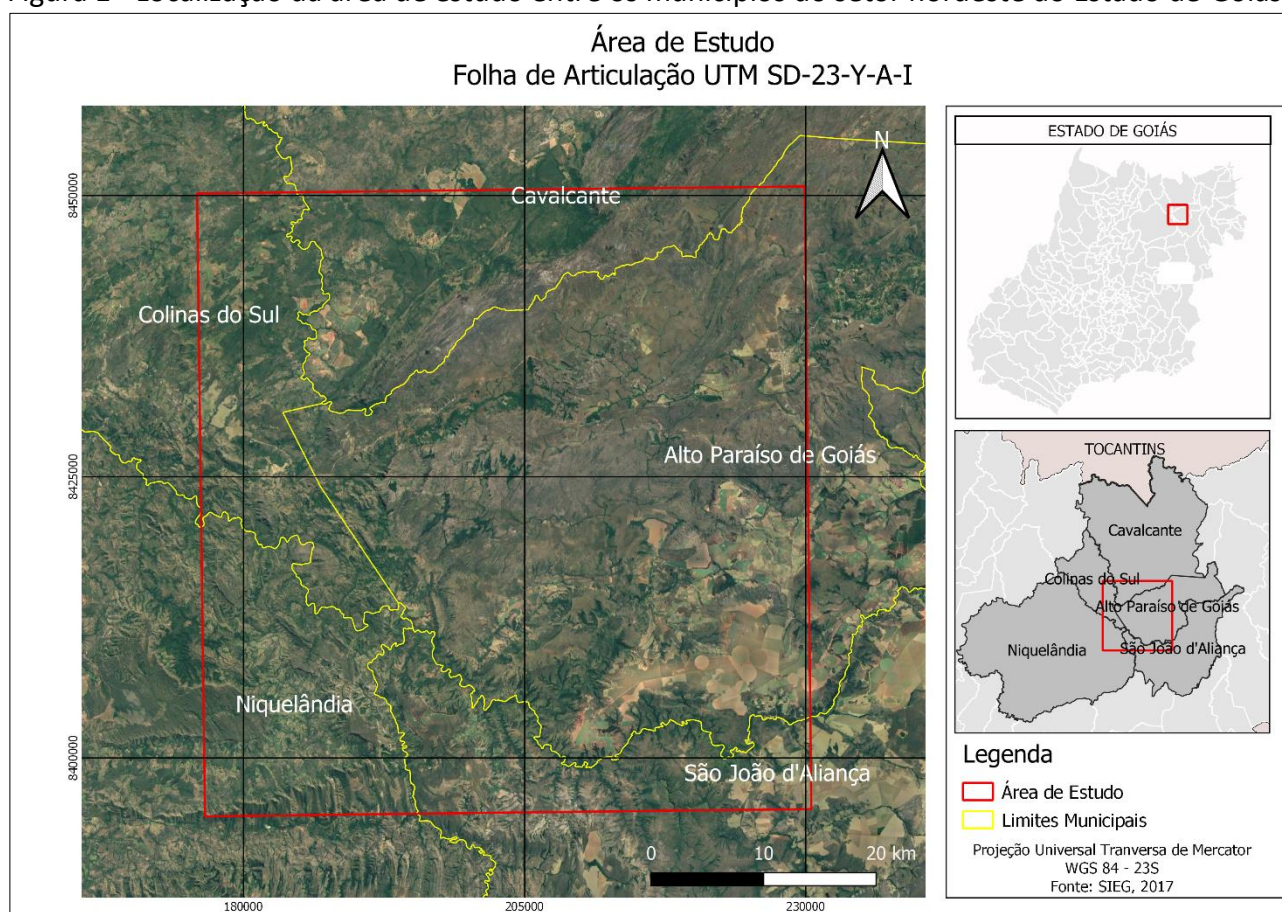
O armazenamento e processamento dessas imagens em disco local demanda infraestrutura computacional. Plataformas baseadas na nuvem, como o Google Earth Engine (GEE), permitem amenizar e contornar limitações de infraestrutura computacional. O GEE é uma plataforma de processamento de dados geoespacial desenvolvida para análises geoespaciais (GORELICK *et al.*, 2017). As capacidades existentes dentro do GEE, tais como uma biblioteca consolidada de imagens de satélite gratuitas, possibilitam o monitoramento contínuo da superfície terrestre por sensores com diferentes resoluções espectrais, espaciais e temporais, inclusive para a integração entre dados ópticos e de radar (GHORBANIAN *et al.*, 2020). Sendo assim, o objetivo deste trabalho foi avaliar a integração de imagens ópticas e SAR a partir da plataforma GEE para a classificação supervisionada da cobertura da terra de uma região contida no bioma Cerrado.

METODOLOGIA

ÁREA EXPERIMENTAL

Como área experimental para o desenvolvimento deste trabalho foi selecionada uma área representativa dos aspectos fitofisionômicos e de uso antrópico (agropecuária) do bioma Cerrado. A área selecionada (quadrícula de articulação SD-23-Y-A-I) possui área aproximada de 3.230 km² e perímetro de 230 km (Figura 1). Essa área abrange os municípios de Alto Paraíso de Goiás, Cavalcante, Colina do Sul, Niquelândia e São João d’Aliança localizados na região Nordeste do estado de Goiás, na microrregião da Chapada dos Veadeiros (Figura 1).

Figura 1 - Localização da área de estudo entre os municípios do setor nordeste do Estado de Goiás.



Fonte: Elaborado pelos autores (2021).

MATERIAIS

Para a realização desse estudo foi necessária a aquisição dos seguintes materiais:

- Imagens das missões S-1 e S-2, referentes ao período de janeiro a dezembro de 2019; e
- Mapa de cobertura do solo do MapBiomas (2019), disponível online em: http://mapbiomas.org/pages/database/mapbiomas_collection.

As imagens S-1, no GEE, consistem em cenas com comprimento de onda na banda C, em *Ground Range Detected* (GRD) de nível 1, com as polarizações VH e VV disponíveis e processadas para o coeficiente de retroespalhamento (σ°) em decibéis (dB). O coeficiente de retroespalhamento representa a área de retroespalhamento do alvo (seção transversal do radar) por unidade de área de solo. As imagens adquiridas pela plataforma GEE já possuem: (i) arquivos de órbita aplicados; (ii) remoção de ruído de borda GRD; (iii) remoção de ruído térmico; (iv) calibração radiométrica; e (v) correção do terreno (ortorretificação).

No caso da missão do S-2 foram utilizadas doze bandas espectrais (azul (b2), verde (b3), vermelho (b4), bandas da borda do vermelho (b5, b6, b7, b8A), infravermelho próximo (b8), vapor d'água (b9), cirrus (b10), e infravermelho de ondas curtas 1 (b11) e 2 (b12)). As imagens S-2 foram processadas como nível-1C, o que significa que elas são imagens ortorretificadas e que contêm dados de reflectância no topo da atmosfera (ToA). A opção de trabalhar com imagens em ToA, ao invés de reflectância de superfície foi devido a maior disponibilidade de imagens ToA no catálogo. Foram selecionadas imagens com percentual de cobertura de nuvens menor que 20%. Do mesmo modo, foi empregada a bitmask QA60 (banda de correção de nuvens e cirrus do S-2), para mascarar nuvens opacas e cirros.

MÉTODOS

Os procedimentos metodológicos realizados para a execução do presente estudo são apresentados na Figura 2.

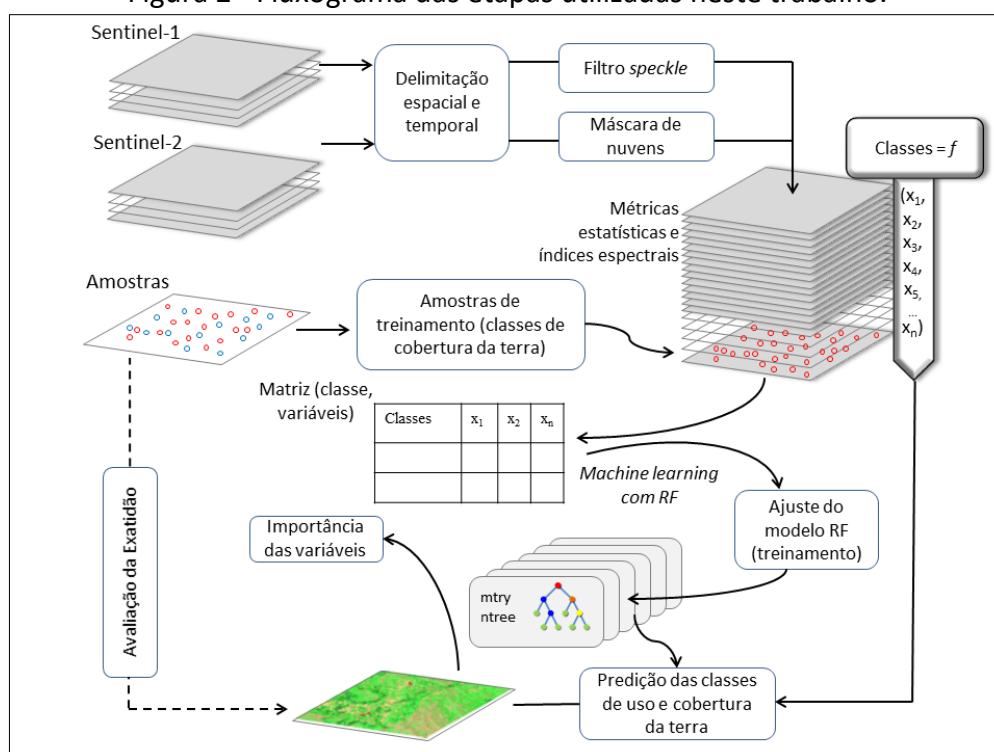
a) *Pré-processamento dos dados SAR e ópticos*

O pré-processamento dos dados SAR consistiu na aplicação de filtro adaptativo para reduzir o efeito *speckle* e atenuar o efeito do ruído que degrada a qualidade da imagem SAR. Segundo

Sant’Anna, Fernandes e Lacava (2001), o ruído dificulta a interpretação visual dos dados, assim como a classificação destas imagens. Por isso, foi necessária a redução deste ruído para obter uma melhor análise visual. Para tanto, foi utilizado o filtro *Refined Lee* (LEE et al., 2008) em janela móvel 7x7.

Para toda a série temporal foram aplicados filtros estatísticos (redutores) temporais em relação aos valores de σ^0 para cada polarização (VH e VV), onde para cada pixel foram calculadas as estatísticas dos valores mínimos, máximos, médios, mediana, variância e o desvio padrão. De forma similar, essas estatísticas também foram derivadas para as bandas individuais e índices espectrais (Quadro 1) extraídos das imagens ópticas da missão do S-2.

Figura 2 - Fluxograma das etapas utilizadas neste trabalho.



Fonte: Elaborado pelos autores (2021).

Quadro 1 - Índices espectrais utilizados no estudo. BLUE = banda com comprimento de onda na faixa do azul; GREEN = verde; RED= vermelho; NIR = infravermelho próximo; REDEDGE = borda do vermelho; SWIR1 = infravermelho de ondas curtas 1; SWIR2 = infravermelho de ondas curtas 2.

Índice	Fórmula
NDVI	$(NIR - RED) / (NIR + RED)$
EVI2	$2,5 * ((NIR - RED) / (NIR + 2,4*RED + 1))$
CAI	$SWIR2 / SWIR1$
NDWI	$(NIR - SWIR1) / (NIR + SWIR1)$
GCVI	$(NIR / GREEN - 1)$
Hall Cover	$(-RED*0,017 - NIR*0.007 - SWIR2*0,079 + 5,22)$
PRI	$(BLUE - GREEN) / (BLUE + GREEN)$
SAVI	$(1 + L) * (NIR - RED) / (NIR + RED + 0,5)$
NDRE	$(NIR-REDEDGE) / (NIR+REDEDGE)$
NDBI	$(SWIR1-NIR) / (SWIR1+NIR)$

Fonte: Souza *et al.* (2020).

b) Amostras de treinamento, classificação e validação

Para a coleta de amostras de treinamento no processo de classificação supervisionada, a seleção das amostras foi realizada com base na interpretação visual de imagens ópticas do S-2 com auxílio secundário do mapa de uso e cobertura do solo do Mapbiomas (MAPBIOMAS, 2019). As amostras coletadas representam 7 (sete) classes de cobertura do solo, as mesmas utilizadas no mapeamento feito pelo Mapbiomas (2019). As classes de cobertura da terra utilizadas e seus respectivos IDs são descritos no Quadro 2.

Como regra geral, Congalton (1991) recomenda pelo menos 50 amostras por classe. Se a área exceder 500 km² ou o número de categorias for superior a 12, pelo menos 75–100 amostras devem ser coletadas por classe. Além disso, a amostragem pode ser alocada com respeito à variabilidade dentro de cada categoria. O número de amostras por classe é apresentado no Quadro 2. A amostragem total foi particionada em amostras de treinamento (70 %) e validação (30%).

Quadro 2 - Classes utilizadas na classificação da cobertura do Cerrado e suas respectivas descrições.

ID	Classe	Número de amostras	Descrição
0	Água	100	Rios, lagos, represas, reservatórios e outros corpos d'água
1	Urbano	100	Áreas urbanizadas com predomínio de superfícies não vegetadas, incluindo estradas, vias e construções
2	Floresta	104	Tipos de vegetação com predomínio de espécies arbóreas
3	Cobertura agrícola	105	Áreas predominantemente ocupadas com cultivos anuais, perenes e semi-perenes
4	Pastagem	100	Áreas de pastagens, naturais ou plantadas, vinculadas à atividade agropecuária
5	Savana	109	Formações savânicas com estratos arbóreo e arbustivo-herbáceos definidos
6	Campo	115	Formações campestres com predominância de estrato herbáceo e algumas áreas de formações savânicas

Fonte: Elaborado pelo autor (2021).

Para a classificação supervisionada no mapeamento da cobertura da terra foi utilizado o classificador Random Forest (RF) que consiste na combinação de várias árvores de decisão preditoras (BREIMAN, 2001). Essas árvores de decisão são criadas de forma independentes através de um subconjunto de amostras de treinamento (BREIMAN, 2001). No GEE, o RF pode ser utilizado a partir da função *ee.Classifier.smileRandomForest* tendo como argumento as amostras de treinamento e os dados matriciais, no caso, as imagens do S-1 e S-2 com os redutores estatísticos temporais. Para o uso do algoritmo RF foi necessária a inserção de alguns hiperparâmetros, como o número de árvores construídas (*ntree*) e o número de variáveis amostradas aleatoriamente como candidatas em cada árvore de decisão (*mtry*).

A classificação pelo algoritmo RF foi feita considerando três cenários: (1) classificação com a integração dos dados ópticos e SAR; (2) classificação utilizando apenas os dados ópticos; e (3) classificação usando apenas os dados SAR. Nos três casos, os mesmos hiperparâmetros foram aplicados. Para o *mtry*, utilizou-se a raiz quadrada do número de variáveis (e.g. métricas estatísticas das bandas, polarizações e índices espectrais), como sugerido por Breiman (2001). Para o *n tree* (número de árvores de decisão) foi escolhido o valor de 500, uma vez que o aumento de árvores com $n > 500$ não apresentou melhora nas classificações conforme testes prévios.

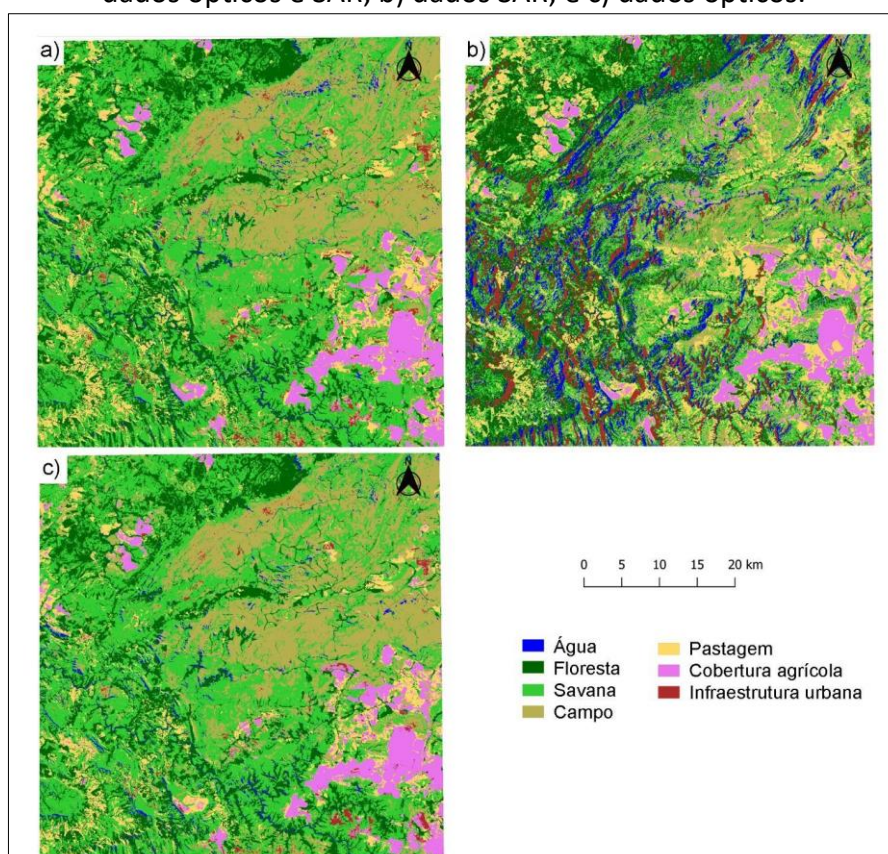
A avaliação da exatidão do classificador RF foi realizada a partir da matriz de confusão. Essa avaliação foi realizada no GEE utilizando o API *ConfusionMatrix*, que possibilitou a tabulação cruzada das classes reais (amostras de validação) e mapeadas (preditas pelo classificador). Com base na matriz de confusão foram calculadas as métricas de índice kappa e exatidão global para os três cenários. Isto foi feito para avaliar se a classificação integrada usando os dados ópticos e radar foi melhor do que as classificações dos demais cenários. Além do melhor cenário, avaliaram-se também quais variáveis foram as mais importantes entre as classificações. Pelo algoritmo RF foi utilizado o índice de Gini para avaliar a quantidade relativa que cada variável contribuiu para a separabilidade entre as classes na classificação supervisionada. Valores mais altos de índice de Gini representam maior importância na classificação final e uma maior contribuição para maximizar a diferença entre as classes. Por fim, foram obtidas as métricas de exatidão do produtor, exatidão do usuário e kappa condicional na avaliação das classes separadamente.

RESULTADOS

Todo o código usado no GEE para este trabalho pode ser acessado pelo endereço: <https://bit.ly/3b6BJUE>. Nos três cenários testados (Figura 3 e 4), a melhor classificação foi integrando os dados ópticos e SAR. A classificação da integração destes dados obteve exatidão global e kappa de 94,12% e 93,07%, respectivamente. Já a classificação das imagens de radar obteve exatidão global e kappa de 79,29% e 75,78%, respectivamente. Por fim, a classificação usando apenas os dados ópticos obteve uma exatidão global de 91,44% e um índice kappa de 89,99%.

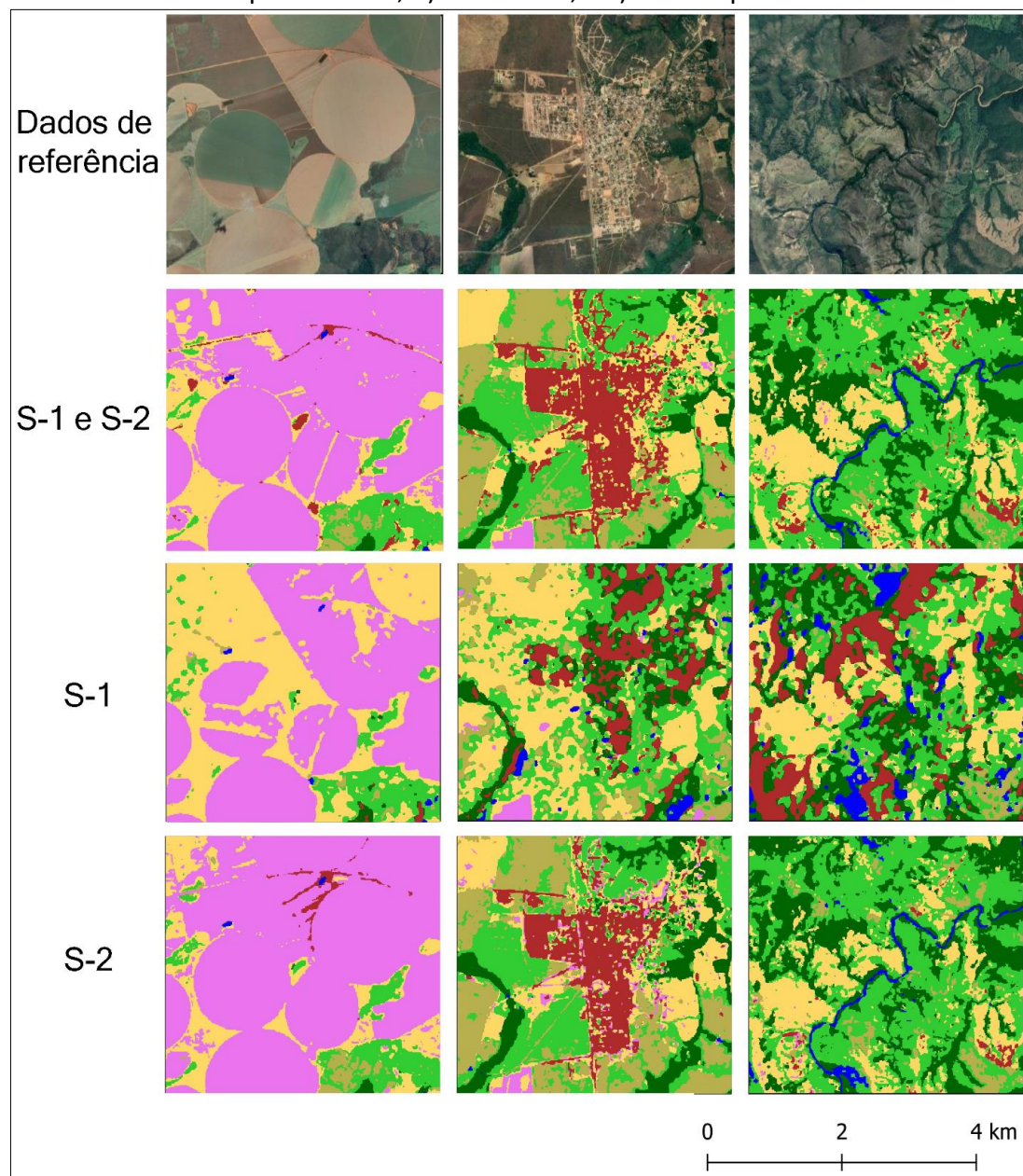
Tendo como base os resultados do kappa condicional por classe (Figura 5), é possível notar que a maioria das classes teve desempenho superior a 90% , sobretudo para as classificações usando o S-2 ou a integração entre S-1 e S-2. O pior desempenho por classe foi para cobertura agrícola e pastagem. Ao analisar a exatidão do usuário, isto é, para verificar o quanto do que foi classificado para uma determinada classe realmente corresponde a essa classe, a pastagem foi a que teve o pior desempenho, seguido da classe savana. Já pela exatidão do produtor, ou seja, que permite avaliar o quanto de uma dada classe foi predita corretamente pelo classificador, o pior desempenho foi para a classe savana, seguido da pastagem (Figura 5).

Figura 3 - Classificação da cobertura da terra na área de estudo com base na: a) integração de dados ópticos e SAR; b) dados SAR; e c) dados ópticos.



Fonte: Elaborado pelos autores (2021).

Figura 4 - Detalhes da comparação entre os três cenários de classificação: a) integração de dados ópticos e SAR; b) dados SAR; e c) dados ópticos.



Fonte: Elaborado pelos autores (2021).

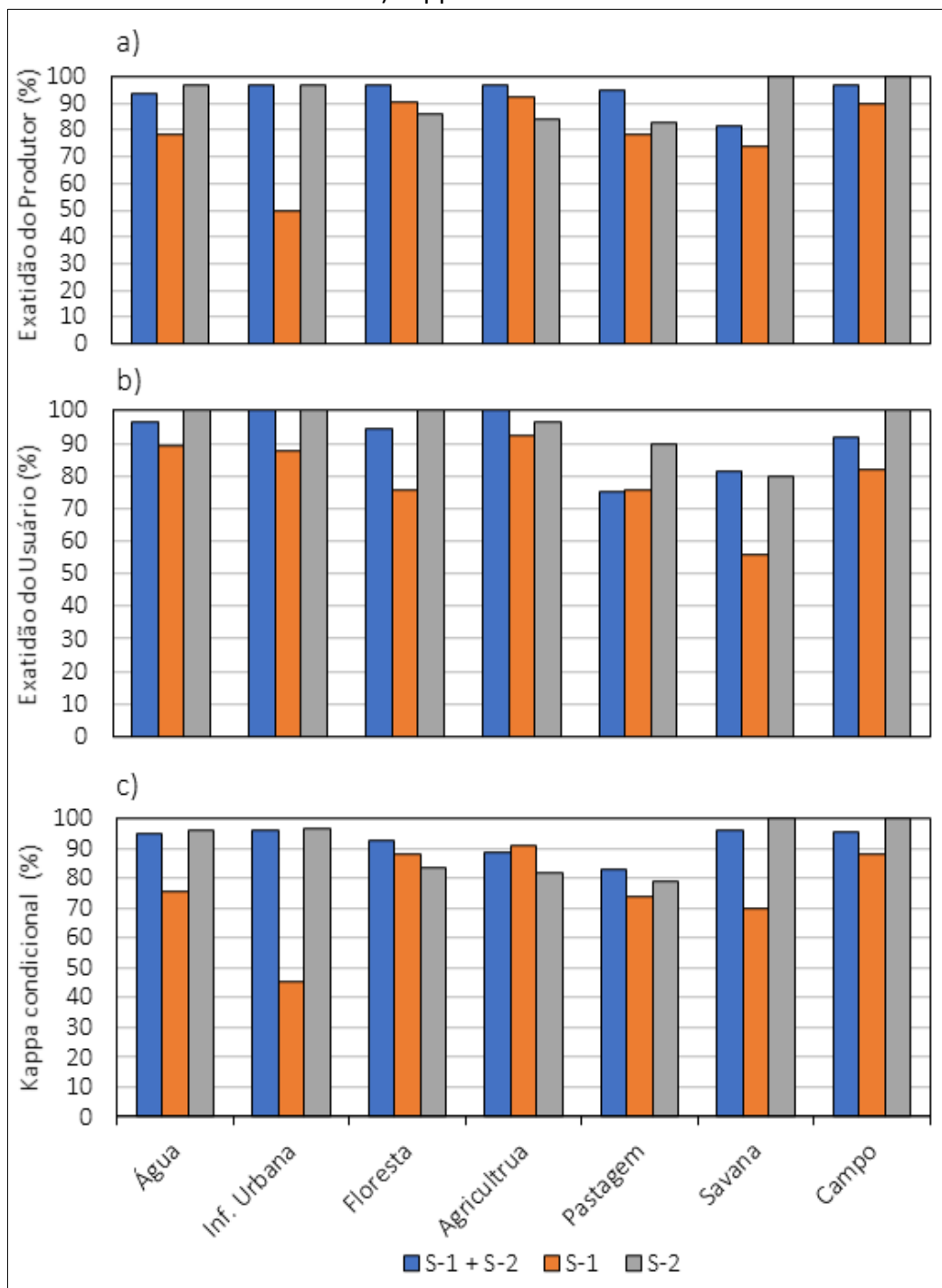
Para todas as métricas avaliadas por classe, no geral, a classificação resultante dos dados SAR (Figura 5), foi a que teve o pior desempenho quando comparado com os cenários de integração

entre dados ópticos e SAR e usando os dados ópticos apenas. Nesta classificação, o melhor desempenho por classe foi para a classe cobertura agrícola.

A classificação com os dados ópticos revelou um desempenho de 100% para as classes água, infraestrutura urbana, floresta e campo, tendo como base a exatidão do usuário (Figura 5). Essas classes apresentaram desempenhos superiores mesmo quando comparadas a classificação pela integração de dados ópticos e SAR. Pela exatidão do produtor e o kappa condicional, as classes floresta, cobertura agrícola e pastagem tiveram desempenhos inferiores à integração dos dados (Figura 5).

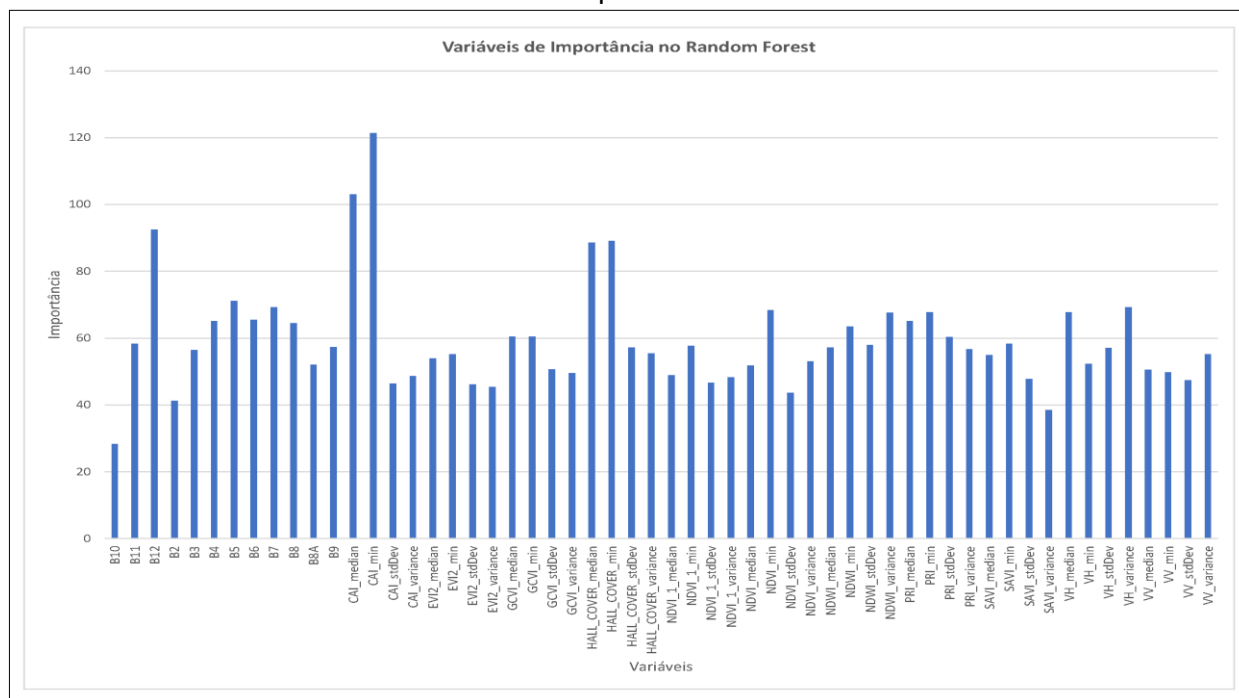
Na classificação com a integração dos dados S-1 e S-2, o índice CAI se mostrou o mais importante, seguido da banda do infravermelho de ondas curtas (SWIR2 - banda 12) e os valores mínimos e medianos do índice Hall Cover (Figura 6). O índice Hall Cover tem em sua formulação as bandas no vermelho, infravermelho e o infravermelho de ondas curtas (Quadro 1).

Figura 5 - Métricas de validação por classe, sendo a) exatidão do produtor; b) exatidão do usuário; e c) kappa condicional.



Fonte: Elaborado pelos autores (2021).

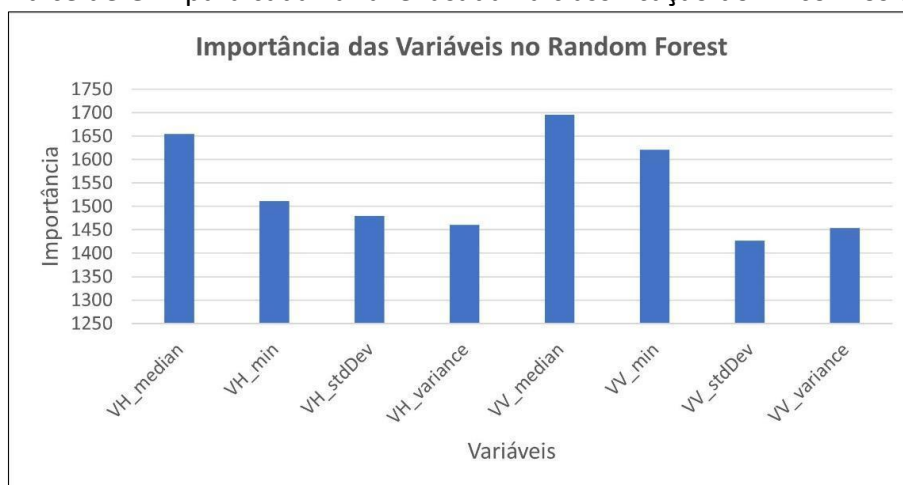
Figura 6 - índice de Gini para cada variável usada na classificação do RF com a integração entre os dados ópticos e SAR.



Fonte: Elaborado pelos autores (2021).

No caso da classificação com os dados SAR, a variável de maior importância foi a polarização VV (mediana; Figura 7), seguida pela polarização VH (mediana).

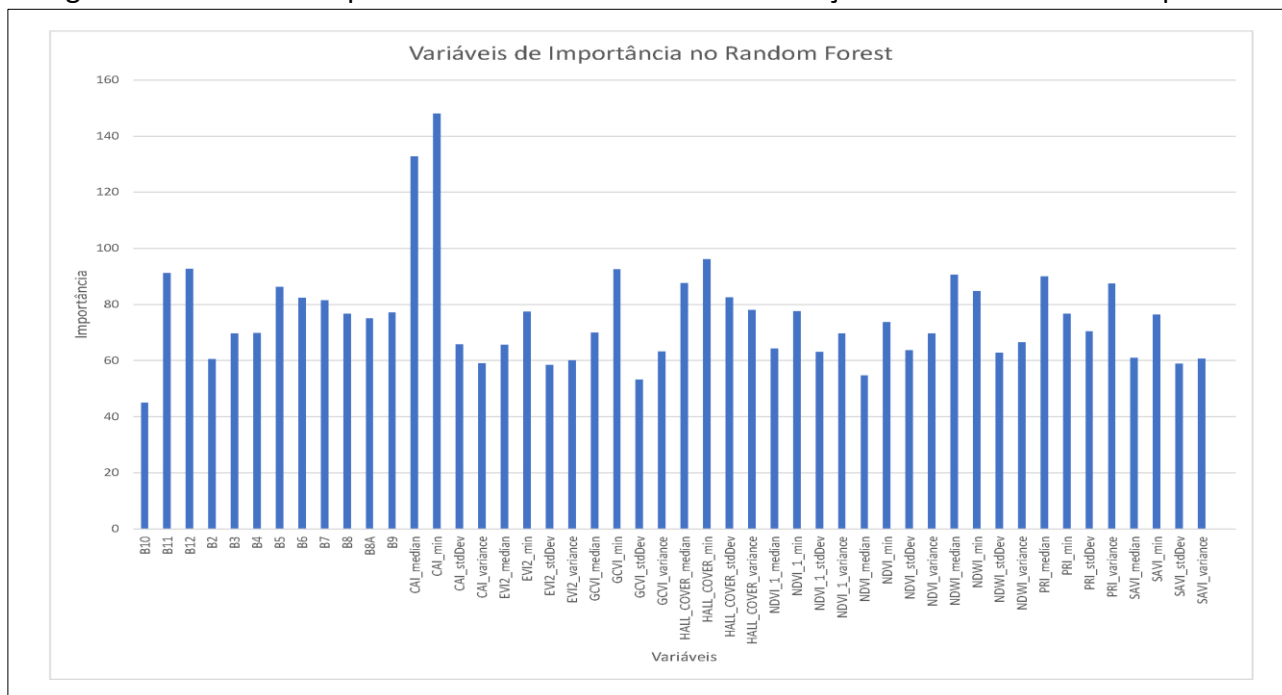
Figura 7. Índice de Gini para cada variável usada na classificação do RF com os dados SAR.



Fonte: Elaborado pelos autores.

Similarmente à classificação utilizando os sensores integrados, a variável mais importante na classificação usando as imagens do sensor óptico foi o índice CAI (valores mínimos e medianos). Secundariamente, o valor mínimo do índice Hall Cover e a banda do infravermelho de ondas curtas também se apresentaram com grande importância na classificação (Figura 8).

Figura 8. Índice de Gini para cada variável usada na classificação do RF com os dados ópticos.



Fonte: Elaborado pelos autores.

DISCUSSÕES

Os resultados mostraram que os cenários que usam apenas os dados do S-1 apresentaram desempenho inferior ao S-2 na maioria das classes. O melhor cenário utilizando o S-2 produziu uma exatidão global de 91,44% em comparação com 79,29% para o melhor modelo utilizando apenas o S-1, uma diferença significativa (teste z com nível de confiança de 95%). Isso corrobora com Lopes et al. (2020) que comparou imagens ópticas com imagens SAR nas regiões subsaarianas e com Borges et al. (2020) que comparou a integração dos dados ópticos e SAR em diferentes estações em uma área de conservação na Tanzânia.

Os resultados também mostraram que a combinação dos dados do S-1 e S-2 alcançaram maior exatidão global em comparação com os cenários de classificação com sensor único. A integração desses dados aumentou a exatidão para as classes de floresta, cobertura agrícola e pastagem. Sano et al. (2010) também obteve maior exatidão na classificação geral onde usou sensores ópticos e radar em regiões de savana e pastagem.

Chatziantoniou, Psomiadis e Petropoulos (2017) ao utilizar a integração de imagens do S-1 e S-2 com o algoritmo *Support Vector Machine* (SVM) para o mapeamento de zonas úmidas no Mediterrâneo obtiveram exatidão global e kappa de 94,82% e 93,62%, respectivamente. Zhang *et al.* (2019) sugeriram a integração de SAR com dados ópticos, uma vez que eles podem corrigir erros, particularmente em áreas altamente produtivas (por exemplo, pântanos de irrigação, campos e gramíneas perenes) que podem ser erroneamente classificados como árvores.

A variável de radar que mais se destacou no cenário de classificação combinando os dados das duas missões do Sentinel (S-1 e S-2) foi a polarização cruzada VH, resultado que está em concordância com o trabalho de Mendes *et al.* (2019) que fez classificações utilizando S-1, S-2, TanDEM-X e ALOS PALSAR-2. Porém, em seu trabalho a polarização cruzada VH que mais se destacou foi a do ALOS-PALSAR2, com banda L que possui comprimento mais longo que o da banda C do S-1.

Em relação a importância das variáveis no processo de classificação, a razão da banda SWIR2 pela SWIR1 pelo índice CAI foi a variável mais importante, tanto pela métrica estatística da mediana, quanto pela mínima. Isso provavelmente ocorreu pelo fato dessa razão de banda no infravermelho de ondas curtas realçar o conteúdo de água na folha da vegetação (PONZONI; SHIMABUKURO; KUPLICH, 2015). Isto é compatível com o fato da vegetação do Cerrado ser caracterizada por sazonalidade em suas características fitofisionômicas nos períodos de estiagem e chuva. Por sua vez, o índice Hall Cover foi outra variável importante na separação de classes com o uso do algoritmo RF. O Hall Cover foi originalmente proposto para estimativa de biomassa acima do solo (HALL *et al.*, 2006) e tem sido utilizado pela equipe do MapBiomass (MAPBIOMASS, 2019) como uma das variáveis no projeto de mapeamento de uso e cobertura da terra no Brasil e também no Chaco na América do Sul.

Individualmente, a banda espectral mais importante foi a banda 12 (infravermelho de ondas curtas). No comprimento de onda dessa banda (~2.190 nm) é possível ter uma nítida separação entre solo exposto e vegetação (PONZONI; SHIMABUKURO; KUPLICH, 2015; NOVO,

2010). Tal característica pode ter contribuído para auxiliar a separação das classes de cobertura da terra.

Já em relação aos dados SAR, a polarização mais importante foi a VH e VV com as métricas estatísticas de mediana e variância. Em estudo efetuado por Saraiva (2015) a polarização VH permitiu melhor discriminação de culturas agrícolas do que a polarização VV. O retroespalhamento volumétrico entre os diferentes alvos analisados, dado pela polarização VH, pode ter contribuído para que essa variável se apresentasse como a mais importante entre os dados SAR.

Mercier *et al.* (2019) ao usarem o S-1 na classificação teve como melhor variável a polarização VV. Esses resultados coadunam-se aos encontrados nesse trabalho onde a polarização VV também se mostrou a mais importante, entretanto contrastam quando os dados ópticos são integrados, em que a polarização VH passa a ser mais importante. Mercier *et al.* (2019) em uma classificação utilizando apenas o S-2 tiveram como variável mais importante a banda 12 (infravermelho de ondas curtas), e ao classificar utilizando a integração dos dados do S-1 e S-2 as variáveis mais importantes foram as variáveis derivadas do sensor óptico, resultados esses que está em conformidade com os resultados da presente pesquisa.

Em relação ao uso do processamento em nuvem, neste estudo foram utilizadas 31 imagens SAR e 130 imagens ópticas (janeiro a dezembro de 2019). Na plataforma Copernicus essas imagens são disponibilizadas para download sem nenhuma ou pouca correlação com tamanho de 1,6 Gb e 500 Mb em formato compactado, respectivamente, totalizando em mais 120 Gb de imagens compactadas. Além do grande consumo da memória de disco, processar esses dados demandaria muito tempo e um alto esforço de processamento da máquina. Kumar e Mutanga (2018) compararam o processamento em nuvem com o processamento em disco local desktop utilizando o trabalho de Hansen *et al.* (2013) que identificaram mudanças na cobertura florestal em escala global durante 12 anos (2000-2012) usando 654.178 cenas do Landsat-7, totalizando em 70 terabytes de dados. O processamento durou 100 horas no GEE em comparação com aproximadamente 1.000.000 horas que teria levado em um computador desktop. Tais números

realçam o potencial e as vantagens da capacidade de armazenamento e processamento em nuvem, como o fornecido pelo GEE.

De acordo com Magidi *et al.* (2021), há uma grande vantagem na combinação da técnica do RF com plataformas de gerenciamento de big data, como o GEE, que é o processamento de grandes conjuntos de dados com inúmeras variáveis em um curto espaço de tempo. Porém, é apontado como desvantagem do algoritmo sua tendência de se ajustar excessivamente aos dados de treinamento (*overfitting*), razão principal para eles não fornecerem uma exatidão ainda melhor. Porém, Oshiro (2013) explica que a técnica de poda evita o problema de *overfitting*, por meio da geração de uma hipótese mais genérica a partir do conjunto de treinamento.

CONSIDERAÇÕES FINAIS

A integração entre dados ópticos do S-2 e dados de radar do S-1 para o mapeamento da cobertura da Terra por classificação supervisionada pelo algoritmo RF em parte do Bioma Cerrado usando a plataforma GEE permitiu concluir que:

- A integração entre os dados ópticos e SAR obteve resultados melhores do que o uso individual desses dados na discriminação da vegetação, áreas de cobertura agrícola, pastagem e áreas de infraestruturas urbanas (classe que apresenta confusão pela mistura espectral de diversos alvos). A integração se mostrou eficaz para mapear o bioma Cerrado, com potencial para o monitoramento da cobertura da Terra;
- A combinação de dados multitemporais e multissensores para mapear regiões heterogêneas, como o Cerrado, sugere que em áreas sujeitas a períodos com alta cobertura de nuvens onde a coleta de dados ópticos não é viável, a integração de dados com diferentes características espectrais, incluindo dados de radar e ópticos tem potencial para um mapeamento confiável e contínuo;
- A banda 12 (infravermelho de ondas curtas), a polarização VH, os índices CAI e Hall Cover foram as variáveis mais importantes na separação das classes propostas de cobertura da terra neste trabalho;

- As métricas estatísticas de valor mínimo, mediana e desvio padrão foram as mais importantes entre as bandas, polarizações e índices espectrais e puderam capturar a sazonalidade espectral dos alvos ao longo do ano permitindo a sua separabilidade entre classes.

REFERÊNCIAS

- BORGES, J. et al. Sentinel-1 and sentinel-2 data for savannah land cover mapping: Optimising the combination of sensors and seasons. **Remote Sensing**, v. 12, n. 23, p. 3862, 2020.
- BRASIL. Ministério do Meio Ambiente. **Mapeamento do uso e cobertura do cerrado: Projeto Terra Class Cerrado 2013**. Brasília, v. 67, 2015.
- BREIMAN, L. Random forests. **Machine learning**, v. 45, n. 1, p. 5–32, 2001.
- CHATZIANTONIOU, A.; PSOMIADIS, E.; PETROPOULOS, G. P. Co-orbital Sentinel 1 and 2 for LULC mapping with emphasis on wetlands in a mediterranean setting based on machine learning. **Remote Sensing**, v. 9, n. 12, p. 1259, 2017.
- CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, v. 37, n. 1, p. 35–46, 1991.
- GHORBANIAN, A. et al. Improved land cover map of Iran using sentinel imagery within google earth engine and a novel automatic workflow for land cover classification using migrated training samples. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 167, p. 276–288, 2020.
- GORELICK, N. et al. Google earth engine: Planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, v. 202, p. 18–27, 2017.
- HALL, R. J. et al. Modeling Forest stand structure attributes using landsat ETM+ data: Application to mapping of aboveground biomass and stand volume. **Forest Ecology and Management**, v. 225, n. 1, p. 378–390, 2006.
- HANSEN, M. C. et al. High-resolution global maps of 21st-century forest cover change. **Science**, v. 342, n. 6160, p. 850–853, 2013.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Mapas de biomas e vegetação**. 2004.

KUMAR, L.; MUTANGA, O. Google Earth Engine applications since inception: Usage, trends, and potential. **Remote Sensing**, v. 10, n. 10, p. 1509, 2018.

LEE, J. S. et al. Improved sigma filter for speckle filtering of sar imagery. **IEEE Transactions on Geoscience and Remote Sensing**, v. 47, n. 1, p. 202–213, 2008.

LOPES, M. et al. Combining optical and radar satellite image time series to map natural vegetation: savannas as an example. **Remote Sensing in Ecology and Conservation**, v. 6, n. 3, p. 316–326, 2020.

MAGIDI, J. et al. Application of the random forest classifier to map irrigated áreas using google earth engine. **Remote Sensing**, v. 13, n. 5, 2021.

MAPBIOMAS, Projeto. **Projeto mapbiomas: Coleção 5.0 da série anual de mapas de cobertura e uso de solo do brasil**. Projeto MapBiomass: Coleção, v. 5, 2019.

MENDES, F. S. et al. Optical and SAR remote sensing synergism for mapping vegetation types in the endangered Cerrado/Amazon ecotone of Nova Mutum—Mato Grosso. **Remote Sensing**, v. 11, n. 10, p. 1161, 2019.

MERCIER, A. et al. Evaluation of Sentinel-1 and 2 time series for land cover classification of forest–agriculture mosaics in temperate and tropical landscapes. **Remote Sensing**, v. 11, n. 8, 2019.

NOVO, E. M. L. M. **Sensoriamento Remoto: princípios e aplicações**. São Paulo: Editora Blucher, 2010.

OSHIRO, T. M. **Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica**. Tese (Doutorado) — Universidade de São Paulo, 2013.

PEREIRA, L. O.; FREITAS C. C.; S., SANT’ANNA S. J.; S., REIS M. ALOS/PALSAR data evaluation for land use and land cover mapping in the amazon region. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 9, n. 12, p. 5413–5423, 2016.

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

SANO, E. E.; JESUS, E. T; BEZERRA, H. S. Mapeamento e quantificação de áreas remanescentes do cerrado através de um Sistema de Informações Geográficas. **Sociedade & Natureza**, v. 13, n. 25, 2001.

SANO, E. E. et al. Land cover mapping of the tropical savanna region in Brazil. **Environmental Monitoring and Assessment**, v. 166, n. 1, p. 113–124, 2010.

SANT'ANNA, S. J. S.; FERNANDES, D.; LACAIVA, J. C. S. Análise comparativa do desempenho de filtros redutores de “speckle” em imagens SAR. In: SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES (SBSR), v. 9, 2001, Fortaleza, **Anais [...]** Fortaleza: SBSR, 2001.

SARAIVA, C. S. A. **Avaliação o Potencial das Imagens Sentinel-1 para Identificação de Culturas Agrícolas**. 2015. Dissertação (Mestrado em Engenharia Geográfica). Universidade de Lisboa, Lisboa, 2015. Disponível em: <https://repositorio.ul.pt/handle/10451/20682>. Acesso: 26 nov. 2020.

SOUZA, C. M. et al. Reconstructing three decades of land use and land cover changes in brazilian biomes with Landsat archive and Earth Engine. **Remote Sensing**, v. 12, n. 17, 2020.

SYMEONAKIS, E. et al. Optimisation of savannah land cover characterization with optical and SAR data. **Remote Sensing**, v. 10, n. 4, 2018.

ZHANG, W. et al. From woody cover to woody canopies: How Sentinel-1 and Sentinel-2 data advance the mapping of woody plants in savannas. **Remote Sensing of Environment**, v. 234, p. 111465, 2019.