

Avaliação de recursos para o mapeamento de cobertura do solo em sub-bacia do rio Capibaribe-PE usando imagem kompsat-2

Juarez Antônio da Silva Júnior, Admilson da Penha Pacheco

Universidade Federal de Pernambuco, e-mail:juarez.silvajunior@ufpe.br (autor correspondente)

<http://orcid.org/0000-0002-2898-0309>

Universidade Federal de Pernambuco, e-mail: admilson.pacheco@ufpe.br <http://orcid.org/0000-0002-3635-827>

Received 25; June 2023, accepted 27 November 2023

RESUMO

O mapeamento de uso e cobertura da terra é um fator importante na análise geoespacial na gestão de bacias hidrográficas. A integração de imagens de sensoriamento remoto e técnicas de classificação por Aprendizagem de Maquinas viabilizam a identificação e monitoramento ambiental dos elementos da paisagem. O sensor MSC (*MultiSpectral Camara*) do satélite Kompsat-2 captura imagens de alta resolução espacial o que permite a identificação de recursos terrestres em escala local. Foram desenvolvidos seis modelos de dados para a classificação do uso e cobertura da terra por *Random Forest* em uma sub-bacia hidrográfica do Rio Capibaribe. Esses modelos foram criados com base em índices espectrais e no ranking de importância de variáveis. A avaliação dos resultados foi feita por meio da quantificação espacial e análise de acurácia. Os produtos baseados em bandas e índices espectrais apresentaram acurácia global variando entre 94 e 98% onde as classes de Vegetação Arbórea e Arbustiva se destacaram com estimativas de acurácia do produtor e usuário acima de 80%. Os produtos com os menores recursos de dados mostraram baixo desempenho de acurácia com valores de acurácia global concentrados abaixo de 60%. Este estudo é o primeiro a usar adaptações de dados espectrais do Kompsat-2 e métodos de aprendizado computacional para demonstrar a aplicação de mapeamento de cobertura do solo de alto desempenho. Dessa forma, este artigo contribuiu para o monitoramento da superfície do solo em sub-bacias hidrográficas urbanas que necessitam de informações espaciais precisas sobre o estado conversação ambiental.

Palavra-chaves: Random Forest, Sensoriamento Remoto, NDVI, Recursos Hídricos.

Assessment of resources for mapping land cover in the Capibaribe-PE river sub-basin using kompsat-2 image

Abstract

Land use and cover mapping is an important factor in geospatial analysis in river basin management. The integration of remote sensing images and Machine Learning classification techniques enable the identification and environmental monitoring of landscape elements. The MSC (*MultiSpectral Camara*) sensor on the Kompsat-2 satellite captures high spatial resolution images, which allows the identification of terrestrial resources on a local scale. Six data models were developed to classify land use and cover by *Random Forest* in a sub-basin of the Capibaribe River. These models were created based on spectral indices and variable importance rankings. The results were evaluated through spatial quantification and accuracy analysis. Products based on bands and spectral indices presented global accuracy ranging between 94 and 98%, where the Tree and Shrub Vegetation classes stood out with producer and user accuracy estimates above 80%. Products with the smallest data resources showed poor accuracy performance with overall accuracy values concentrated below 60%. This study is the first to use adaptations of Kompsat-2 spectral data and computational learning methods to demonstrate the application of high-performance land cover mapping. In this way, this article contributed to the monitoring of the soil surface in urban sub-watersheds that require precise spatial information about the environmental conversation status.

Keywords: Random Forest, Remote Sensing, NDVI, Water Resources.

1. Introdução

As bacias hidrográficas têm uma posição vulnerável em relação a questões associadas a mudanças generalizadas nos padrões de uso e cobertura da terra. Antes da interferência humana, os

sistemas hidrográficos estão em um estado de equilíbrio dinâmico (Alves *et al.* 2021). Nas últimas décadas, a intensificação do manejo da cobertura vegetal em bacias hidrográficas tem alterado a

dinâmica dos sistemas hidrológicos. Modifica as taxas de evaporação, infiltração, erosão e os regimes de superfície e subsuperfície, comprometendo também a disponibilidade e a qualidade da água (Gameiro *et al.* 2022; Tavares *et al.* 2019).

No sensoriamento remoto, os mapas de uso e cobertura do solo são extremamente necessários para o gerenciamento da terra e monitoramento de recursos naturais, como zonas úmidas e secas, áreas florestais, pastagens (AVCI *et al.* 2021). A maneira mais significativa de produzir mapas uso e cobertura do solo é por meio de classificação de imagens de sensoriamento remoto usando algoritmos de classificação (NEDD *et al.* 2021). Os mapas de uso e cobertura do solo são usados em diferentes campos, como rastreamento agrícola, dinâmica espacial em bacias hidrográficas e planejamento urbano, são extremamente importantes no monitoramento sustentável e na análise diferentes ecossistemas (AVCI *et al.* 2021; Lu *et al.* 2013). Mudanças no uso e cobertura da terra podem afetar a biodiversidade, ciclos biogeoquímicos de carbono e nitrogênio, balanço de energia, escoamento de água, entre outros (Tavares *et al.* 2019). Também poderia levar a uma redução na prestação de serviços ecossistêmicos (AVCI *et al.* 2021; NEDD *et al.* 2021; Alves *et al.* 2021). Para a mudança de cobertura do solo a longo prazo análise geral, as imagens de satélite Landsat foram usadas por fornecer imagens de satélite gratuitas desde a década de 1970 (Hussain *et al.* 2022). No entanto, as imagens de satélite Kompsat-2 ganharam importância na análise de mudanças de curto prazo (Sunwoo *et al.* 2016). Com capacidade para 4 bandas espectrais, alta resolução espacial entre 1-4 metros e acesso gratuito em escala global, o Kompsat-2/MSR oferece uma grande vantagem para análises ambientais.

Estudos têm sido feitos para simplificar o processo de mapeamento de uso e cobertura do solo com o auxílio de índices espectrais (Chaves *et al.* 2023). Um índice espectral pode ser calculado usando duas ou mais bandas de comprimento de onda das imagens de satélite detectadas remotamente. O Índice de Vegetação de Diferença Normalizada (NDVI), o Índice de Vegetação Aprimorado (EVI) e o Índice de Água de Diferença Normalizada (NDWI) são três índices frequentemente usados. O NDVI descreve abundância de vegetação e zonas de estresse

vegetativo (Ponzoni; Shimabukuro; Kuplich, 2015). Gao (1996) e McFeeters (1996) propuseram um Índice de Água de Diferença Normalizada (NDWI) para sensoriamento remoto da abundância de água usando as bandas de comprimento de onda verde visível e infravermelho próximo, enquanto EVI monitora de vegetação por meio de um desacoplamento do sinal de fundo do dossel e uma redução nas influências da atmosfera.

Muitos métodos de classificação diferentes foram desenvolvidos e conduzidos, como métodos de classificação não supervisionados e supervisionados (MA *et al.* 2017). Com o desenvolvimento de algoritmos de *Machine Learning* (ML), a precisão e o uso de métodos de classificação aumentaram. Existem vários algoritmos de *Machine Learning* usados para a classificação de uso e cobertura do solo, mas o algoritmo de floresta aleatória (RF) tornou proeminente nos últimos anos e têm sido amplamente utilizados para esse fim (AVCI *et al.* 2021; GHOSH *et al.* 2014). Esse método fornece um desempenho superior em comparação aos métodos tradicionais, especialmente em aplicações de sensoriamento remoto (Lu *et al.* 2013). No entanto, esse algoritmo não fornece os mesmos resultados com diferentes dados de satélite. A qualidade, número e distribuição dos conjuntos de dados de treinamento e teste são de grande importância para o desempenho do algoritmo selecionado. É importante comparar aplicativos para medir o desempenho desses algoritmos com base no número de pontos de treinamento e teste (Silva Junior *et al.* 2022).

Este estudo tem como objetivo analisar a influência dos recursos de dados de sensoriamento remoto na qualidade da classificação de uso e cobertura do solo numa sub-bacia no município de Vertentes-PE. Para isso, foi utilizado uma cena do sensor MSR (*MultiSpectral Camera*) do satélite Kompsat-2 bem como índices espectrais NDVI (*Normalized Difference Vegetation Index*), NDWI (*Normalized Difference Water Index*) e EVI (*Enhanced Vegetation Index*) e o classificador Random Forest por meio da informação de importância da variável na aplicação do modelo. Dessa forma, foi gerado seis produtos de dados com recursos espectrais diferentes possibilitando a análise de acurácia da classificação para cada produto de uso e cobertura do solo.

2. Material e métodos

2.1 Área de estudo

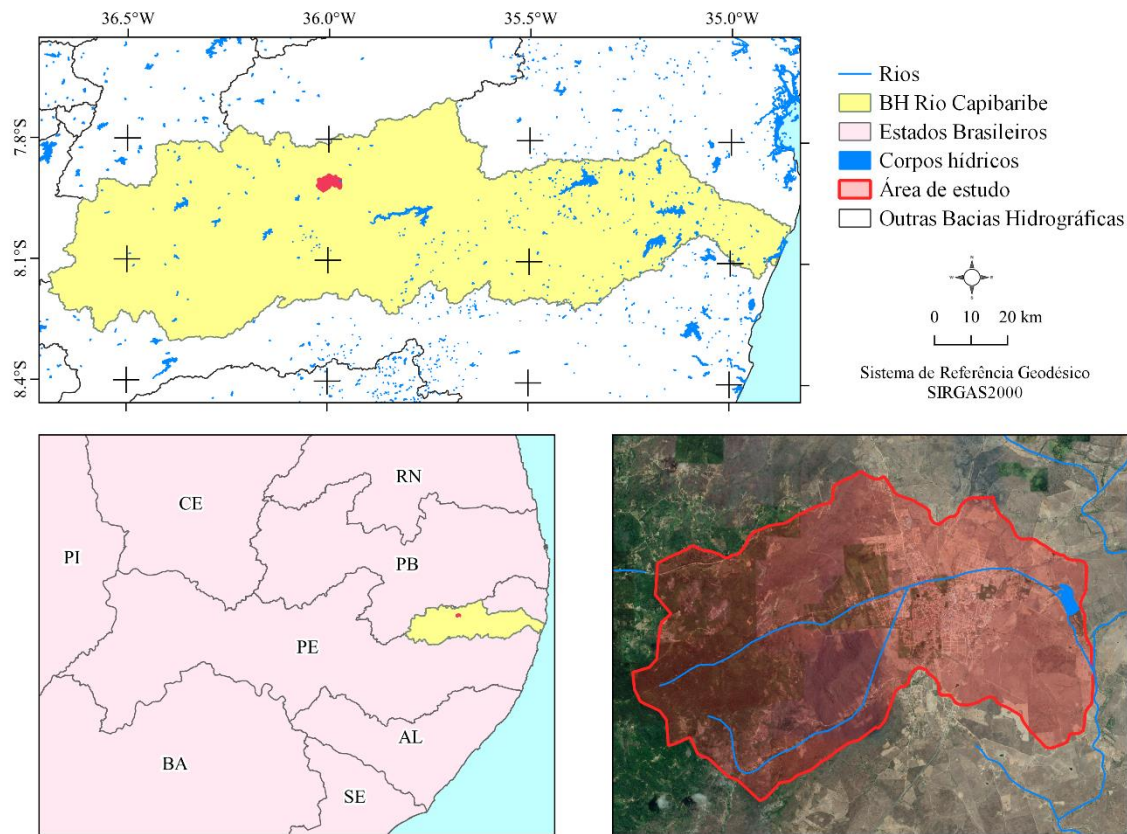
A área de estudo abrange uma sub-bacia hidrográfica do Rio Capibaribe (Fig 1), localizada no

estado de Pernambuco, mais precisamente no município de Vertentes. Essa sub-bacia possui uma

área de 21,67 km² e está completamente inserida nesse município.

A região é caracterizada por um clima semiárido, o qual é marcado por chuvas escassas e irregulares ao longo do ano. A estação chuvosa ocorre de janeiro a maio, sendo curta e intensa, enquanto a estação seca se estende de junho a novembro, sendo

longa e rigorosa. Essas características climáticas são típicas dessa região, que enfrenta condições adversas de disponibilidade hídrica, o que pode ter impactos significativos na gestão dos recursos hídricos e no desenvolvimento sustentável da área (Assis; Lacerda; Sobral, 2012).



A vegetação nesta área é predominantemente caracterizada como caatinga, é possível encontrar espécies de árvores como a jurema-preta (*Mimosa tenuiflora*) e baraúna (*Schinopsis brasiliensis*). Também são encontradas várias espécies de cactos, bromélias, xique-xique (*Cereus jamacaru*) e outras plantas adaptadas às condições do clima semiárido (OLIVEIRA *et al.* 2018).

O Kompsat-2/MS (Korean Multi-purpose Satellite 2), é um satélite coreano de imagem da Terra/ambiente que foi lançado em 28 de julho de 2006 e permanece operacional, porém sem registro de imagens. O sensor MSC (*Multi-Spectral Camera*) tinha como finalidade capturar imagens pancromáticas e multiespectrais da superfície da Terra (ESA, 2008). Os parâmetros de desempenho do MSC são resumidos na tabela 1.

2.2 Dados

Tabela 1.Principais especificações Técnicas da cena KOMPSAT-2/MS

Parâmetro	Valor
Alcance Espectral	500-900 nm (PAN), 450 - 520 nm (Azul), 520 - 600 nm (Verde), 630 - 690 nm (Vermelho), 760 - 900 nm (NIR)
Resolução espacial	Pan (1 m), MS (4 m)
Largura da Faixa	15 km no nadir, ±0,62°
Resolução radiométrica	10 Bits

Fonte: (ESA, 2008)

A cena MSC foi capturada na data 08/08/2010 e obtido pelo catalogo de imagens da Agência Nacional Europeia (ESA) disponível em: <https://earth.esa.int/eogateway/catalog> onde recobre integralmente a área de estudo sem a presença de nuvens. As imagens Kompsat-2 são disponibilizadas na escala de pixels em ND (Número Digital), diante disso, para a representação física dos alvos terrestres

2.3 Índices Espectrais

Após a obtenção bandas em níveis de refletância, três índices de vegetação diferentes foram calculados para esta pesquisa e comumente preferidos

na imagem, foi necessária a conversão em ND para reflectância ToA (*Top of Atmospheric*). Para atingir este objetivo, foram utilizadas as informações de metadados e os métodos descritos em Kim *et al.* (2015), que leva em consideração a radiância espectral média da banda, ângulo zênite solar e a distância terra-sol. No fim do processamento, foram obtidas as bandas de reflectância ToA nas faixas multiespectrais.

na literatura, tais o NDVI, NDWI e EV. As especificações dos índices são apresentadas na Tabela 2.

Tabela 2. Especificações dos índices espectrais usados

Índice	Abreviação	Formula	Referência
Normalized Difference Vegetation Index	NDVI	$\frac{NIR - Red}{NIR + Red}$	Rouse <i>et al.</i> (1974)
Normalized Difference Water Index	NDWI	$\frac{NIR - Red}{NIR + Red}$	Mcfeeters (1996)
Enhanced Vegetation Index	EVI	$G * \frac{NIR - Red}{NIR + C1 * Red - C2 * Blue + L}$	Huete <i>et al.</i> (1997)

O NDVI utiliza a luz vermelha e infravermelha refletida pela vegetação. Os cálculos de NDVI para um determinado pixel sempre resultam em um número que varia de menos um (-1) a mais um (+1). No entanto, a falta de folhas verdes faz com que o valor fique próximo de zero. Um valor zero indica ausência de vegetação e um valor próximo a +1 (0,8–0,9) indica a maior densidade possível de folhas verdes (Sheykhmousa *et al.* 2020).

O índice de vegetação aprimorado (EVI) é um índice de vegetação 'otimizado' projetado para melhorar o sinal de vegetação com sensibilidade aprimorada em regiões de alta biomassa e monitoramento de vegetação aprimorado por meio de um desacoplamento do sinal de fundo do dossel e uma redução nas influências da atmosfera (ZHANG *et al.* 2014). Na formulação do EVI, $G = 2,5$ e $L = 1$ representam o ganho e fundo do dossel, respectivamente, enquanto $C1 = 6$ e $C2 = 7,5$ representam os coeficientes usados para corrigir as condições atmosféricas (Huete *et al.* 1997). O NDWI é utilizado para destacar as características da água e o teor de umidade, obtido pelo cálculo de combinação GREEN-NIR (verde visível e infravermelho próximo), o que lhe permite detectar mudanças sutis no teor de água das massas de água (Mcfeeters, 1996).

2.4 Seleção de amostras

A seleção de amostras para a classificação de imagem é uma etapa importante pois definem as classes consideradas na classificação e são os principais dados de entrada para o treinamento do modelo. Neste artigo foi adotado o método de seleção amostral sugerido por Körting *et al.* (2014) que descreve de coleta de amostras deve procurar dentro das classes características com um pequeno desvio padrão, calculando a homogeneidade intraclasse. Dessa forma, A seleção de amostras foi realizada de forma aleatória e estratificada por meio de interpretação visual baseados em polígonos vetoriais categorizados para cada classe. Com a finalidade de reduzir inconstâncias espaciais e temáticas, foi utilizada um recorte do produto Mapbiomas com a identificação das classes de cobertura do solo em relação ao reconhecimento de padrões na imagem fusionada Kompsat-2 em cores naturais.

2.5 Classificador Random Forest e Importância de Gini

O classificador de floresta aleatória consiste em uma combinação de classificadores de árvore onde cada classificador é gerado usando um vetor aleatório amostrado independentemente do vetor de entrada, e cada árvore lança um voto unitário para a classe mais popular para classificar um vetor de entrada (Breiman, 2001). O treinamento e classificação por RF foi

aplicado através da combinação de produtos espectrais e as amostras de treinamento, trazendo como resultado, um modelo raster categorizados com as classes de uso e ocupação do solo utilizando o software SNAP.

Para implementar o RF, dois parâmetros precisam ser configurados: o número de árvores (*ntree*) e o número de recursos em cada divisão (*mtry*). Este estudo seguiu a recomendação de Noi e Kappas (2017) que afirmaram que o modelo RF implementado para classificação da cobertura do solo alcança valores de acurácia estáveis quando *ntree* = 500. Com relação ao parâmetro *mtry*, foi utilizado o valor padrão *mtry* = $\sqrt{p}$, onde *p* é o número de variáveis predatórias (Noi e Kappas, 2017; Breiman, 2001).

Após o processamento do classificador RF é gerado a proporcionalidade importância de predição de cada variável no modelo gerado. O Score Gini calcula a importância de cada recurso como a soma do número de divisões (em todas as árvores) que incluem o recurso, proporcionalmente ao número de amostras que ele divide (Archer; Kimes, 2008). Matematicamente, o Índice de Gini pode ser expresso como:

$$Gini\ index = 1 - \sum_{i=1}^n (p_i)^2 \quad (1)$$

Onde p_i denota a probabilidade de um elemento ser classificado para uma classe distinta. A diminuição média no coeficiente de Gini é uma medida de como cada variável contribui para a homogeneidade dos nós e folhas na floresta aleatória resultante (Behnamian *et al.* 2017). Quanto maior o valor do escore de Gini, maior a importância da variável no modelo (Breiman, 2001).

Baseados nos resultados de importância, as análises foram definidas através 6 diferentes produtos de dados denominados (P1-P6), selecionadas com

base no nível de importância de cada variável. Dessa forma, inicialmente, o produto P1 é composta por todos os recursos espectrais (Azul, Verde, Vermelho, IVP, NDVI, NDWI e EVI), os demais produtos serão selecionados através na análise de ranking no Score de Gini.

2.6 Validação do uso e cobertura da terra

A avaliação da confiabilidade da classificação da cobertura do solo foi obtida por meio da Acurácia Global (AG) eq (2). O AG é calculado dividindo-se a soma da diagonal principal da matriz de contingência x_{ii} pelo número total de amostras coletadas *n* :

$$Acurácia\ Global = \frac{\sum_{i=1}^c x_{ii}}{n} \quad (2)$$

Onde x_{ii} é o valor na linha *i* é a coluna *i* e *c* é o número de classes. A precisão de uma classe foi representada pela precisão do produtor e do usuário. A precisão do produtor refere-se a amostras que não foram classificadas corretamente como pertencentes a essa categoria e omitidas de sua categoria correta, e a precisão do usuário indica a probabilidade de que um pixel classificado na imagem realmente represente essa categoria (Congalton *et al.* 2009).

Dessa forma, os parâmetros de acurácia são obtidos pelos elementos da tabela de contingência gerados pela comparação do produto classificado e dado de referência. Uma cena UAV do Programa Pernambuco Digital (PE3D) foi adotada como referência recobrindo integralmente a área de estudo. O PE3D é o mapeamento aerofotogramétrico de todo território do Estado de Pernambuco realizado entre 2010 e 2011, que disponibiliza ortofotos georreferenciados na escala 1:5.000, resolução espacial de 1m, atendendo ao mais alto nível de qualidade cartográfica. As imagens estão disponíveis gratuitamente em <http://www.pe3d.pe.gov.br/>

3. Resultados

3.1 Análise de Importância da Variável e definição dos produtos de dados

Após a realização do modelo de combinação P1 e a análise das amostras de treinamento, foi observado no gráfico da figura 5 que o NDVI teve ligeira vantagem como a variável mais importante no modelo,

seguido pela banda IVP e NDWI. O EVI e a banda vermelha tiveram pontuações moderadas, com o EVI tendo uma pequena vantagem. Os scores mais baixos foram atribuídos à banda verde e azul, respectivamente.

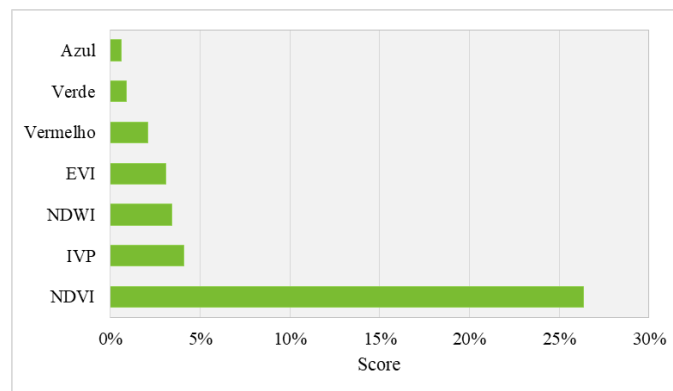


Figura 1. Gráfico de barras com a informação de score para a importância da variável para cada recurso de dados utilizado.

Diante disso, com base nesses resultados em níveis de ranking de importância, os produtos de dados espectrais utilizados no classificador RF foram selecionados da seguinte forma: P1 = Azul, Verde, Vermelho, IVP, NDVI, NDWI e EVI; P2 = Azul,

Verde, Vermelho, IVP, NDWI e EVI; P3 = Azul, Verde, Vermelho, NDWI e EVI; P4 = Azul, Verde, Vermelho, EVI; P5 = Azul, Verde, Vermelho; P6 = Azul, Verde.

3.2 Mapeamento de cobertura do solo

Segundo Santos et al. (2021) as áreas das classes de cobertura do solo nas bacias hidrográficas podem variar muito, dependendo das características da bacia hidrográfica da localização específica. Esta informação pode ser usada para informar as decisões de gestão da terra e para monitorar as mudanças no uso da terra ao longo do tempo. Dessa forma, a quantificação de

áreas de uso e cobertura do solo em bacias hidrográficas é importante para entender as mudanças na paisagem e como elas afetam o ciclo hidrológico e a qualidade da água em uma determinada área. A figura 3 mostra a distribuição espacial das classes de cobertura do solo obtidas pelo classificador RF para produto de dados.

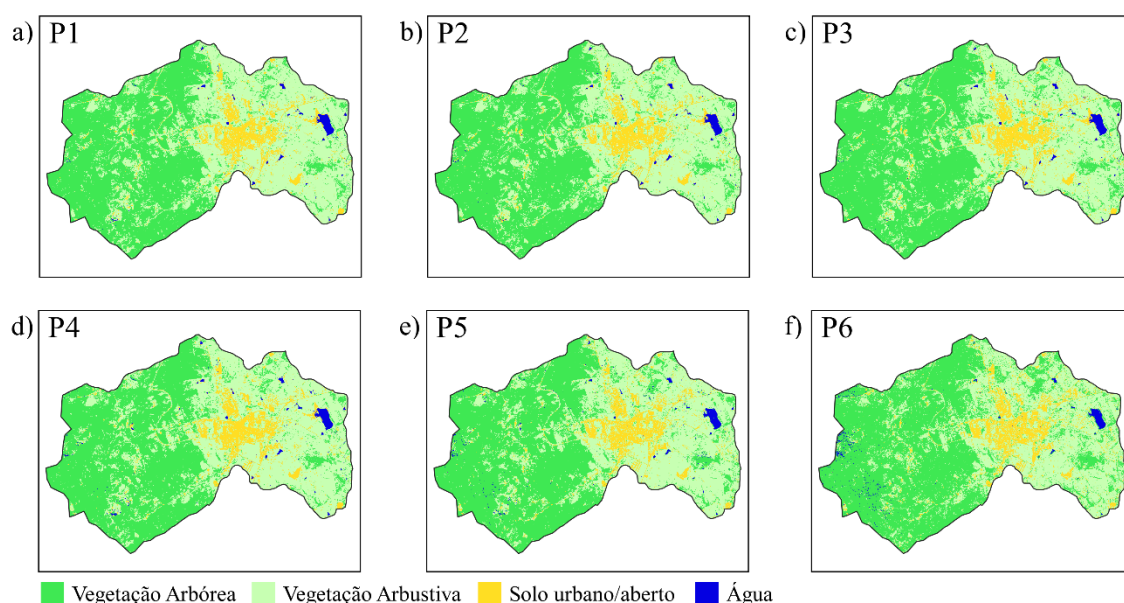


Figura 3. Distribuição espacial da cobertura do solo através do classificador RF na sub-bacia do Rio Capibaribe.

Observa-se na figura que os produtos não apresentaram variações visuais significativas, onde a

maior proporção espacial foi para a classe de vegetação arbustiva, representando aproximadamente

uma variação 47%50 % em relação a área total da bacia.

O produto P1 mostrou o melhor rendimento na comparação entre áreas com o produto de validação, com uma dispersão média de 0,05%, enquanto os produtos P2, P3, P4, P5 e P6 apresentaram uma dispersão de 4,03%, 7,31%, 7,85%, 8,82% e 13,58%. As classes de Solo exposto/urbano e Água para o produto P1 não apresentaram variações em relação ao produto de referência, por outro lado os demais produtos apresentaram uma dispersão variando entre 6~17,28 Km² e 0,06~0,53 Km² respectivamente.

Com redução dos recursos espectrais do NDVI, ou seja, de P1 para P2, houve uma dispersão de 2,29% para a classe de vegetação arbórea, de P2 para P3, caracterizado pela ausência da banda NIR a dispersão foi de 6,37%. Esses resultados mostram a importância desses recursos para o mapeamento de

cobertura do solo em especial para as classes de vegetação densas. Comportamento semelhante foi visto para na comparação do P3 para P4 com a ausência do NDWI para a classe de Água, a dispersão passou de 3,86% para 18,77%, caracterizando o NDWI como um recurso essencial para a qualidade do mapeamento de águas superficiais.

Os produtos P5 e P6, consideradas apenas pelas bandas Azul, Verde e Vermelho mostraram o menor rendimento, principalmente na separação entre vegetação arbustiva e solo exposto/urbano, com uma superestimação significativa chegando a aproximadamente 20,27% e que também vista na classe de Água com uma dispersão de 13,58 para P6 e 12% e 8,82% para P5 respectivamente. A tabela 3 resume as estimativas de área, para todos os produtos classificados.

Tabela 3. Área para as classes de cobertura do solo para os produtos P1-P6.

Classes	P1 (Km ²)	P2 (Km ²)	P3 (Km ²)	P4 (Km ²)	P5 (Km ²)	P6 (Km ²)
Vegetação Arbórea	164,15	168,13	174,83	167,65	174,95	185,43
Vegetação Arbustiva	155,39	149,48	142,21	149,78	146,65	138,2
Solo exposto/urbano	24,81	26,79	27,41	26,53	27,79	29,84
Água	2,82	2,76	2,71	3,35	3,13	3,1

No geral, os valores obtidos na tabela 4 concordam com a distribuição da cobertura do solo em sub- bacias hidrográficas do Alto Capibaribe, onde as florestas arbóreas e arbustivas são dominantes. As áreas não vegetadas caracterizadas como urbanas são concentradas no entorno das sedes municipais, o que facilita a etapa de detecção por imagens de satélite. Por outro lado, são espectralmente semelhantes com as áreas de solo exposto e de pastagens de solo aberto, causando confusões no algoritmo classificador, intensificadas também pela alta resolução espacial que aumenta os recursos de informações. Na área de estudo não possui grandes reservatórios de água.

Os corpos hídricos possuem uma área superficial concentradas abaixo de 3Km² e bastante distribuídas em toda bacia. A partir no produto P4 áreas tenderam a aumentar de extensão, principalmente pela má detecção de sombras por relevo acidentado, onde os recursos de dados com base apenas nas bandas espectrais unitárias, acarretaram em áreas da classe Água superestimadas.

3.3 Análise de acurácia

A análise de acurácia da classificação de uso e cobertura do solo é uma etapa essencial para avaliar a qualidade do mapa gerado. É através dessa análise que se pode avaliar a precisão da classificação em relação aos dados de referência, o que garante a confiabilidade dos resultados e a qualidade da classificação de uso e cobertura do solo (Talukdar *et al.* 2020). Para avaliar a acurácia da classificação, é importante considerar também fatores como a resolução espacial da imagem, a complexidade da paisagem e as condições climáticas.

Neste estudo, a imagem Kompsat foi capturada no período seco, onde a vegetação no geral, encontra-se mais esparsa, solo exposto com feições realçadas e corpos hídricos com características de reduções do espelho d'água (Chughtai *et al.* 2021). Com a diminuição dos recursos entre os produtos, produziram variações na detecção de mudanças consistentes entre classes como mostrado na figura 2.

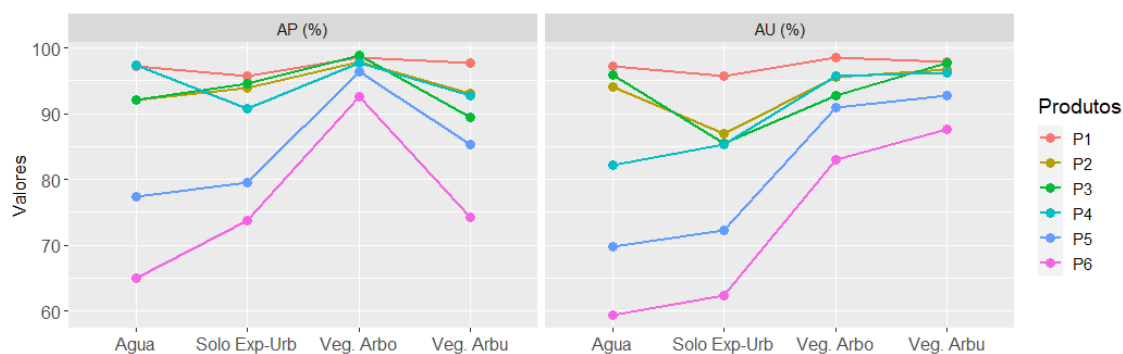


Figura 4. Gráfico de linhas com os valores de acurácia do produtor (AP%) e acurácia do Usuário para os produtos P1-P6.

É perceptível que o produto P1 possui os melhores valores e mais estáveis de acurácia, para todas as classes com valores acima de 95% indicando que a integração do NDVI com outros recursos espectrais aumenta a variação extraclasse do modelo, mesmo para classes de vegetação distantes entre si. Baeza e Paruelo (2020) revelaram que o NDVI permitiu a distinção significativa entre diferentes tipos de vegetação e outros tipos de uso do solo, como áreas urbanas e corpos d'água mostrando alto rendimento nas análises de acurácia no mapeamento de mudanças espaço-temporais de uso e ocupação do solo na América do Sul.

Os produtos P2, P3 e P4 mostraram um declínio perante as classes, porém com valores de acurácia concentrados acima de 80%. Os produtos P2 e P3 mostraram similaridades nos valores de acurácia do produto e usuário, indicando com a que a ausência do NDVI e NIR pode não representar variações significativas entre classes do uso do solo entre si. Apesar disso, o NIR é fortemente refletido pela vegetação saudável bem como a sua densidade, permitindo alta capacidade de detecção (Thenkabail *et al.* 2005).

Houve uma diminuição significativa no produto P3 e P4 na acurácia do usuário nas classes de solo Exp/urbano e Água, provavelmente causado por

pixels escuros, sem uma identificação espectral bem definida pelo algoritmo, pode ter causado essas superestimações como mencionado anteriormente. A retirada do NDWI resultante no produto P4 pode ter intensificado essas inconsistências, principalmente relacionados a classe de Água.

Os produtos P5 e P6 baseados são baseados exclusivamente nas bandas do visível. Os baixos valores de acurácia para os produtos P5 e P6 é ligeiramente perceptível. As classes de solo exposto/urbano e Água foram as mais afetadas. Para esses dois produtos, em todas as classes em geral, a diminuição dos valores de acurácia do produtor e do usuário de P5 para P6 foi de aproximadamente 8%. A classe de Vegetação Arbustiva obteve valores consistentes em todos os produtos, indicando baixa sensibilidade na diminuição de recursos dados.

Segundo (Machado, 2020) as bandas do visível podem ter dificuldade em distinguir entre objetos de cores semelhantes ou que possuem texturas e informações espectrais semelhantes, tornando difícil a distinção entre diferentes tipos de cobertura do solo. Além disso, as bandas do visível também são influenciadas por fatores atmosféricos, como nuvens, neblina e poeira, que podem afetar a qualidade e a precisão dos dados coletados (Forkuor *et al.* 2017). A Tabela 4 mostra as estimativas de acurácia Global.

Tabela 4. Acurácia Global (%) para os produtos P1-P6.

Produtos Classificados	Acurácia Global (%)
P1	97,91
P2	95,33
P3	94,98
P4	94,14
P5	55,01
P6	50,62

As estimativas de acurácia global destacaram os produtos P1-P4 como as melhores combinações de recursos de entrada para a

classificação da cobertura do solo, enquanto os produtos P5 e P6 mostraram uma abrupta diminuição neste parâmetro, causado pela ausência do índice EVI.

Apesar do EVI ser sensível a refletância do solo e a geometria do sensor, o índice é menos afetado por fatores atmosféricos, como por exemplo neblina, que é fortemente ocorrido no Agreste Pernambucano, o que o torna mais preciso em áreas com maior presença de elementos de cobertura do solo (Oliveira *et al.* 2018; Souza *et al.* 2022).

Por fim, classificação de objetos por RF em imagens de Kompsat-2 apresenta vários desafios devido à natureza complexa dos dados. Uma delas é que a cena MSC de resolução espacial de 4m contém uma grande quantidade de dados, dificultando o armazenamento, o processamento e a análise, além uma adicional complexidade com tendência a levar à redução da precisão da classificação devido ao ruído e outros fatores (Noi e Kappas, 2017; GHOSH *et al.* 2015). Por parte do RF existe o risco de *overfitting* do modelo. Isso ocorre quando o modelo aprende muito bem os dados de treinamento, com a mudança dos conjuntos de dados, pode dificultar a determinação de qual objeto é o assunto de interesse e levar a limitações na interpretabilidade (Breiman, 2001).

5 Referências

- Archer, Kellie J.; Kimes, Ryan V, 2008. Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis*, [S.L.], v. 52, p. 2249-2260, jan.. Elsevier BV.
<http://dx.doi.org/10.1016/j.csda.2007.08.015>
- AVCI, Cengiz; Budak, Muhammed; Yağmur, Nur; Balçık, Filiz, 2021. Comparison Between Random Forest and Support Vector Machine Algorithms for LULC Classification. *International Journal Of Engineering And Geosciences*, [S.L.], v. 8, p. 1-10, 30 nov.. International Journal of Engineering and Geoscience.
<http://dx.doi.org/10.26833/ijeg.987605>
- Alves, W. S; Martins, A. P; Pôssa, E. M; Moura, D. M. B. Morais, W. A.; Ferreira, R. S, Santos, L. N. S, 2021. Geotechnologies applied in the analysis of land use and land cover (LULC) transition in a hydrographic basin in the Brazilian Cerrado. *Remote Sensing Applications: Society and Environment*, [S.L.], v. 22, p. 100495, abr. 2021. Elsevier BV.
<http://dx.doi.org/10.1016/j.rsase.2021.100495>
- Assis, J. Lacerda, F. F.; Sobral, M. C, 2012 Análise de Detecção De Tendências no Padrão Pluviométrico na Bacia Hidrográfica do Rio Capibaribe. *Revista Brasileira de Geografia Física*, [S.L.], v. 5, p. 320, 11 out. 2012. *Revista Brasileira de Geografia*

4. Conclusão

No geral, demonstrou que as imagens de satélite Kompsat-2 são uma ferramenta valiosa para a análise de classificação da cobertura da terra e seu uso levou a melhorias significativas na precisão e eficiência do mapeamento e monitoramento em bacias hidrográficas. O Random Forest foi útil no treinamento dos dados e classificação, fornecendo a importância da variável nos diferentes produtos sendo essenciais para a discretização do desempenho dos mapas com base na variabilidade de dos recursos espectrais.

O NDVI e o NIR foram as variáveis que melhor descreveram o modelo de classificação de uso e cobertura do solo e com o alcance de altos valores de acurácia para as classes de Vegetação, por outro lado, o NDWI, foi imperdível na qualidade da detecção dos corpos hídricos. O modelo baseado em bandas espectrais somente na faixa do visível mostraram subestimações significativas. Finalmente, os resultados obtidos podem ser usados para examinar o impacto do aumento das áreas florestais na biodiversidade, incluindo os efeitos sobre as espécies animais e vegetais além de informar o planejamento e gestão sustentável do uso da terra.

- Física.
<http://dx.doi.org/10.26848/rbgf.v5i2.232796>
- Breiman, L, 2001. Random Forests. *Machine Learning*. 45, 5–32 2001.
<https://doi.org/10.1023/A:1010933404324>
- Behnamian, Amir; Millard, Koreen; Banks, Sarah N.; White, Lori; Richardson, Murray; PASHER, Jon, 2017. A Systematic Approach for Variable Selection With Random Forests: achieving stable variable importance values. *Ieee Geoscience And Remote Sensing Letters*, [S.L.], v. 14, p. 1988-1992, nov. 2017. Institute of Electrical and Electronics Engineers (IEEE).
<http://dx.doi.org/10.1109/lgrs.2017.2745049>
- Congalton, R. G., Green, K., Group, F., & Raton, B. 2009. Assessing the accuracy of remotely sensed data—Principles and practices. *International Journal of Applied Earth Observation and Geoinformation*, 11, 448–449.
<https://doi.org/10.1016/j.jag.2009.07.002>
- Cunha, E. R; Santos, C. A. G; Silva, R. M. Bacani, V. M ; Teodoro, P. E. Panachuki, E ; Oliveira, N. S, 2020. Mapping LULC types in the Cerrado-Atlantic Forest ecotone region using a Landsat time series and object-based image approach: a case study of the prata river basin, mato grosso do sul, brazil. *Environmental Monitoring And Assessment*, [S.L.], v. 192, p. 136-146, 24 jan.

2020. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s10661-020-8093-9>
- Chaves, M E. D.; Soares, A. R.; Mataveli, G, A. V.; Sánchez, A, H.; Sanches, I, D. A, 2023 Semi-Automated Workflow for LULC Mapping via Sentinel-2 Data Cubes and Spectral Indices. Automation, [S.L.], v. 4, p. 94-109, 23. MDPI AG. <http://dx.doi.org/10.3390/automation4010007>
- Chughtai, Ali Hassan; Abbasi, Habibullah; Karas, Ismail Rakip. 2021. A review on change detection method and accuracy assessment for land use land cover. Remote Sensing Applications: Society and Environment, [S.L.], v. 22, p. 100482, abr. 2021. Elsevier BV. <http://dx.doi.org/10.1016/j.rsase.2021.100482>
- ESA, European Space Agency. KOMPSAT-2 Image data Manual. 2008.
- Forkuor, Gerald; Dimobe, Kangbeni; SERME, Idriss; Tondoh, Jerome, 2017. Ebagnerin. Landsat-8 vs. Sentinel-2: examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in burkina faso. Giscience & Remote Sensing, [S.L.], v. 55, p. 331-354, 31 ago. 2017. Informa UK Limited. <http://dx.doi.org/10.1080/15481603.2017.1370169>
- Ghosh, A; Sharma, R; Joshi, P.K 2014. Random forest classification of urban landscape using Landsat archive and ancillary data: combining seasonal maps with decision level fusion. Applied Geography, [S.L.], v. 48, p. 31-41, mar. 2014. Elsevier BV. <http://dx.doi.org/10.1016/j.apgeog.2014.01.003>
- Gao, Bo-Cai. 1996 NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing Of Environment, [S.L.], v. 58, p. 257-266, dez. 1996. Elsevier BV. [http://dx.doi.org/10.1016/s0034-4257\(96\)00067-3](http://dx.doi.org/10.1016/s0034-4257(96)00067-3)
- Gameiro, S; Nascimento, V; Facco, D; Sfredo, G; Ometto, J, 2022. Multitemporal Spatial Analysis of Land Use and Land Cover Changes in the Lower Jaguaribe Hydrographic Sub-Basin, Ceará, Northeast Brazil. Land, [S.L.], v. 11, p. 103, 8 jan. 2022. MDPI AG. <http://dx.doi.org/10.3390/land11010103>
- Hussain, Sajjad; Mubeen, Muhammad; Karuppannan, Shankar, 2022. Land use and land cover (LULC) change analysis using TM, ETM+ and OLI Landsat images in district of Okara, Punjab, Pakistan. Physics And Chemistry Of The Earth, Parts A/B/C, [S.L.], v. 126, p. 103117, jun. 2022. Elsevier BV. <http://dx.doi.org/10.1016/j.pce.2022.103117>
- Huete.1997 A comparison of vegetation indices over a global set of TM images for EOS-MODIS. Remote Sensing Of Environment, [S.L.], v. 59, p. 440-451, mar. 1997. Elsevier BV. [http://dx.doi.org/10.1016/s0034-4257\(96\)00112-5](http://dx.doi.org/10.1016/s0034-4257(96)00112-5)
- KIM, Jinsoo; JIN, Cheonggil; CHOI, Chuluong; AHN, Hoyong. Radiometric characterization and validation for the KOMPSAT-3 sensor. Remote Sensing Letters, [S.L.], v. 6, p. 529-538, 10 jun. 2015. Informa UK Limited. <http://dx.doi.org/10.1080/2150704x.2015.1054043>
- Körting, Thales; Fonseca, Leila; Castejon, Emiliano; Namikawa, Laercio, 2014. Improvements in Sample Selection Methods for Image Classification. Remote Sensing, [S.L.], v. 6, p. 7580-7591, 15 ago. 2014. MDPI AG. <http://dx.doi.org/10.3390/rs6087580>
- LU, Dengsheng; LI, Guiying; MORAN, Emilio; HETRICK, Scott. Spatiotemporal analysis of land-use and land-cover change in the Brazilian Amazon. International Journal of Remote Sensing, [S.L.], v. 34, n. 16, p. 5953-5978, 28 maio 2013. Informa UK Limited. <http://dx.doi.org/10.1080/01431161.2013.802825>
- MA, Lei; LI, Manchun; MA, Xiaoxue; CHENG, Liang; DU, Peijun; Liu, Yongxue, 2017. A review of supervised object-based land-cover image classification. ISPRS Journal Of Photogrammetry And Remote Sensing, [S.L.], v. 130, p. 277-293, ago. 2017. Elsevier BV. <http://dx.doi.org/10.1016/j.isprsjprs.2017.06.001>
- Mcfeters, S. K, 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal Of Remote Sensing, [S.L.], v. 17, p. 1425-1432, maio 1996. Informa UK Limited. <http://dx.doi.org/10.1080/01431169608948714>
- Machado, A. M. B, 2020. Comparação de imagens OLI/Landsat-8 e MSI/Sentinel-2 no mapeamento de uso e cobertura da terra na ilha do Maranhão. Revista Equador, [S.L.], v. 9, p. 209-232, 2 jul. 2020. Universidade Federal do Piauí. <http://dx.doi.org/10.26694/equador.v9i3.10451>
- Noi, Phan Thanh; Kappas, Martin, 2017 Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors, [S.L.], v. 18, p. 18, 22 dez. 2017. MDPI AG. <http://dx.doi.org/10.3390/s18010018>
- Nedd, R; Light, K; Owens, Marcia; James, Neil; Johnson, Elijah; Anandhi, Aavudai. 2021, A Synthesis of Land Use/Land Cover Studies: definitions, classification systems, meta-studies, challenges and knowledge gaps on a global landscape. Land, [S.L.], v. 10, p. 994, 21 set. 2021. MDPI AG. <http://dx.doi.org/10.3390/land10090994>

- Oliveira, N.; Maciel, M.; Lúcia, A.; Ulisses A. B.; C, 2018. A. Uso e cobertura do solo utilizando geoprocessamento em municípios do Agreste de Pernambuco. *Revista Brasileira de Meio Ambiente*, v. 4, 2018. Disponível em: <https://revistabrasileirademeioambiente.com/index.php/RVBMA/article/view/95/96>
- Ponzoni, F. J; Shimabukuro, Y. E; Kuplich, T. M, 2015. Sensoriamento remoto da vegetação. *Oficina de Textos*, 2015.
- Rouse, J. W. et al, 1974. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ*, v. 351, p. 309, 1974. <https://ntrs.nasa.gov/citations/19740022614>
- Sunwoo, W.; Kim, D.; Kang, S.; Choi, M, 2016. Application of KOMSAT-2 Imageries for Change Detection of Land use and Land Cover in the West Coasts of the Korean Peninsula. *Korean Journal of Remote Sensing*, v. 32, p. 141–153, 2016. <https://doi.org/10.7780/kjrs.2016.32.2.7>
- Silva J. J. A. Pacheco, A. P; Ruiz-A. A. M; Henriques, R. F. F, 2022 Evaluation of the Ability of SLSTR (Sentinel-3B) and MODIS (Terra) Images to Detect Burned Areas Using Spatial-Temporal Attributes and SVM Classification. *Forests*, [S.L.], v. 14, p. 32, 24 dez. 2022. MDPI AG. <http://dx.doi.org/10.3390/f14010032>
- Souza, B. S; Cantalice, W. L. B. C; Braga, C. C; Silva, M. T; Silva, E. A, 2022. Resposta do índice de vegetação evi a anomalias de precipitação no estado da Paraíba. *Journal Of Hyperspectral Remote Sensing*, [S.L.], v. 12, p. 10, 25 maio 2022. Universidade Federal de Pernambuco. <http://dx.doi.org/10.29150/2237-2202.2022.252412>.
- Sheykhmousa, Mohammadreza; Mahdianpari, Masoud; Ghanbari, Hamid; Mohammadimanesh, Fariba; Ghamisi, Pedram; Homayouni, 2020, Saeid. Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: a meta-analysis and systematic review. *Ieee Journal Of Selected Topics In Applied Earth Observations And Remote Sensing*, [S.L.], v. 13, p. 6308-6325. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/jstars.2020.3026724>
- Santos, J. Y. G; Montenegro, S. M. G. L; Silva, R. M; Santos, C. A. G; Quinn, N. Wyndham; Dantas, A. P. Xavier; R. N, A, 2021 Modeling the impacts of future LULC and climate change on runoff and sediment yield in a strategic basin in the Caatinga/Atlantic forest ecotone of Brazil. *Catena*, [S.L.], v. 203, p. 105308, ago. 2021. Elsevier BV. <http://dx.doi.org/10.1016/j.catena.2021.105308>
- Talukdar, S; Singha, P; Mahato, S; Shahfahad; Pal, Swades; Liou, Y; Rahman, A, 2020. Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. *Remote Sensing*, [S.L.], v. 12, p. 1135, 2 abr. 2020. MDPI AG. <http://dx.doi.org/10.3390/rs12071135>
- Thenkabail, P S; Schull, M; Turrall, H, 2005 Ganges and Indus river basin land use/land cover (LULC) and irrigated area mapping using continuous streams of MODIS data. *Remote Sensing Of Environment*, [S.L.], v. 95, p. 317-341, abr. 2005. Elsevier BV. <http://dx.doi.org/10.1016/j.rse.2004.12.018>
- Tavares, P; Beltrão, N; Guimarães, U; Teodoro, A, 2019. Integration of Sentinel-1 and Sentinel-2 for Classification and LULC Mapping in the Urban Area of Belém, Eastern Brazilian Amazon. *Sensors*, [S.L.], v. 19, p. 1140, 6 mar. 2019. MDPI AG. <http://dx.doi.org/10.3390/s19051140>
- Zhang, X; Tan, B; Yu, Y, 2014 Interannual variations and trends in global land surface phenology derived from enhanced vegetation index during 1982–2010. *International Journal Of Biometeorology*, [S.L.], v. 58, p. 547-564, 18 mar. 2014. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s00484-014-0802-z>