EFEITO DA CLASSIFICAÇÃO SEMIAUTOMATICA COM USO DE SÉRIES TEMPORAIS EM ÁREAS DO BIOMA MATA ATLÂNTICA

Raquel Z. Molinez[1], Rodrigo S. Carmo[1], Deise S. A. Santana[1], Felipe O. Passos[1], Mauricio B. Trancho[1], Janaina Maia[1], Luciana Soler[1], Silvana Amaral[1]

[1]Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos – SP, Brasil,
raquel.zozimo@gmail.com; rod19.silva@gmail.com; deisealvesw@gmail.com; felipeo.passos@gmail.com;
mauriciotrancho@gmail.com; jan.maia07@gmail.com; lusoler@gmail.com; silvana.amaral@inpe.br

RESUMO

Este estudo propõe uma metodologia semiautomática de classificação de séries temporais para monitorar a supressão de vegetação nativa no Prodes Mata Atlântica (MA). Foram analisadas três áreas com intensa ocorrência de nuvens e desmatamentos utilizando imagens Sentinel-2 de 2023, com 10 m de resolução, obtidas pelo *Brazil Data Cube* (BDC). O algoritmo *Random Forest* foi utilizado para o treinamento e classificação de 3.163 pontos amostrais, alcançando acurácia geral entre 79% e 95%, e estatística por classe de até 99%. Apesar da dificuldade em separar classes com semelhanças espectrais, a validação cruzada por k-fold demonstrou média superior a 0,77 em áreas com alta cobertura de nuvens, resultando em mapas de probabilidade com níveis de precisão de 76%, 87% e 86%. O estudo aponta a necessidade de refinamentos futuros no aprimoramento da separabilidade entre classes com padrões espectro-temporais similares.

Palavras-chave **—** machine learning, monitoramento ambiental, Prodes, análise espectral, desmatamento.

ABSTRACT

This study proposes a semiautomatic time series classification methodology to monitor the suppression of native vegetation in the Prodes Mata Atlântica (MA). Three areas with high cloud occurrence and deforestation were analyzed using 2023 Sentinel-2 images, with 10 m resolution, obtained from Brazil Data Cube (BDC). The Random Forest algorithm was employed for the training and classification of 3,163 sample points, achieving overall accuracy between 79% and 95%, and statistics by class up to 99%. Despite the challenges in separating spectrally similar classes, k-fold cross-validation demonstrated an average sample above 0.77 in areas with high cloud cover, resulting in probability maps with precision levels of 76%, 87%, and 86%. The study highlights the need for future refinements, particularly in improving the separability between classes with similar spectral-temporal patterns.

Key words **—** *Machine learning, environmental monitoring, Prodes, spectral analysis, deforestation.*

1. INTRODUÇÃO

A Mata Atlântica (MA) está entre os 35 locais no mundo onde a alta biodiversidade resiste, apesar a grande perda de habitats [1, 2]. A cobertura remanescente de 11,6%, intensamente fragmentada [3], torna-o um dos três hotspots globais de biodiversidade mais vulneráveis às mudanças climáticas [4].

Em 2023, foram suprimidos 765,17 km² de vegetação na MA, 7.812,17 km² na Amazônia e 11.011,69 km² no Cerrado. Embora haja uma tendência de queda desde 2009, esses números refletem que 71% das florestas primárias MA já foi convertida para outro uso [5], indicando a importância da restauração florestal, que pode levar décadas [6].

O sistema Prodes, parte do programa de monitoramento da vegetação brasileira – BiomasBR do Instituto de Pesquisa (INPE) monitora os efeitos sobre a cobertura da terra desde 1989 [7]. O sistema vem sendo ampliado para todos os biomas e aprimorado para obter dados mais precisos sobre o desmatamento [8], possibilitando analises referente aos desmatamentos da MA [9]. O Prodes deverá inovar sua metodologia com o uso de séries temporais e algoritmos de *machine learning*, à medida que seus resultados garantirem a acurácia, a qualidade e a manutenção de toda a série histórica.

No entanto, tem sido um desafio monitorar o acelerado uso dos recursos naturais em larga escala na MA. Ainda mais se tratando do bioma responsável por aproximadamente 70% do PIB nacional no qual habitam 72% da população brasileira [10]. Portanto, são necessárias pesquisas e o constante refinamento das metodologias.

A disseminação gratuita de imagem de satélite possibilitou o uso de séries temporais para gerar informações mais precisas. Assim, avanços nos algoritmos de *machine learning* para treinamento de modelos e classificações já são uma realidade [11]. O potencial destas tecnologias oferece alternativas promissoras para o aprimoramento do monitoramento do desmatamento.

O pacote SITS (*Satellite Image Time Series*) possibilita analise em séries temporais de imagens de satélite, com algoritmos de classificação de *machine learning* e gera classificações de uso e cobertura da terra [12]. No entanto, seu uso na MA apresenta desafios devido à complexidade dos ecossistemas em um bioma de latitudes continentais [13].

Este trabalho apresenta os resultados da fase de treinamento e do modelo de classificação referentes à uma proposta de automatização do monitoramento da supressão

de vegetação nativa no bioma MA. Propõe-se um modelo de classificação semiautomática para identificar a supressão de vegetação para o ano de 2023, fazendo uso de séries temporais bem como algoritmos de *machine learning*.

2. MATERIAL E MÉTODOS

Como área de estudo, foram escolhidas três áreas teste (Figura 1), localizadas nos estados de Pernambuco e Alagoas (Norte – tile 040015), Minas Gerais (Centro – tile 035023), e Santa Catarina e Rio Grande do Sul (Sul – tile 025035). Com 11.151,36 km², dimensão correspondente aos tiles *small* do projeto *Brazil Data Cube* (BDC) [14] estas áreas foram escolhidas por suas variabilidades latitudinais, transição com outros biomas, fitofisionomias, persistência e densidade de polígonos PRODES [9] e taxas de cobertura de nuvem [15].

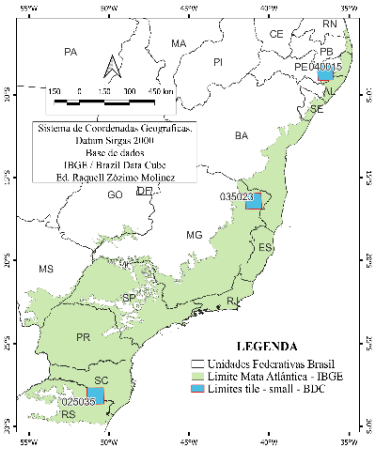


Figura 1. Localização das amostras no bioma MA.

Pontos amostrais foram coletados manualmente, em 8 imagens mosaicos Sentinel-2, de 2023 (duas por trimestre), com resolução de 10 m e composição R8G11B4. A estrutura hierárquica das classes, conforme descrito na Figura 2, propõe diferenciar coberturas naturais e antrópicas. Além dos grupos de classes florestal e não florestal, fez-se a distinção das classes de desmatamento: com solo exposto, com vegetação, e em áreas de não floresta (espécies herbáceas e arbustivas) para os diferentes períodos (trimestres) ao longo do ano. O objetivo foi de capturar os diferentes padrões espaço-temporais possivelmente associados à ocorrência de desmatamento,

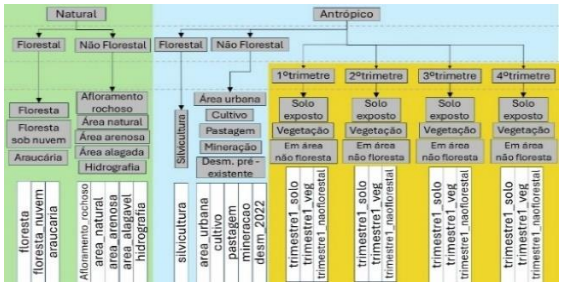


Figura 2. Fluxograma de níveis de avaliação de classes.

Os procedimentos propostos são observados na Figura 3, e descrito a seguir. A produção dos mapas de ocorrência das classes foi realizada no *Rstudio*, com o uso do pacote SITS [12]. O software QGIS foi utilizado para aquisição de amostras e visualização das imagens.

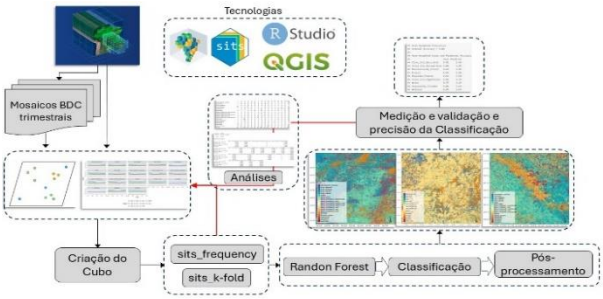


Figura 3. Etapas metodológicas propostas para classificação.

Utilizando a coleção Sentinel-2 16D foram criados os cubos de dados e escolhidos os índices e bandas que melhor diferenciaram as classes contidas nas amostras. Utilizou-se a função *Agglomerative Hierarchical Clustering* - AHC (sits_cluster_dendro) para calcular a dissimilaridade entre amostras e agrupá-las em *clusters*. Foi utilizada a métrica de DTW (*Dynamic Time Wraping*), recomendada para cálculos em séries temporais. Para distinção entre classes foi analisada uma tabela de frequência que organiza as amostras baseada na similaridade espectral para agrupá-las em *clusters* e a validação cruzada, para a medida de estimativa de erro. A metodologia também permite a edição dos *clusters* à medida que se identificam confusões entre classes [16].

O treinamento do modelo foi realizado utilizando o método *Randon Forest* seguido da classificação do cubo gerando os mapas classificados (mapas de probabilidade). Os métodos de pós processamento incluíram a suavização, remoção de *outliers* e rotulagem.

Para as medidas de acurácia dos mapas de probabilidade, foi utilizada a amostragem aleatória estratificada na abordagem *ad-hoc*, na qual gera automaticamente um conjunto de pontos utilizando a função do SITS sits_stratified_sampling. Esses pontos (5.324) foram visualmente analisados no Qgis, e utilizados na acurácia do mapa classificado pela técnica ponderada por área. Esta técnica corrige vieses de erro de mapeamento e inclui intervalos de confiança para evitar superestimações em classes sub-representadas, como áreas de incremento de desmatamento [17, 18].

3. RESULTADOS E DISCUSSÕES

3.1. Análise das amostras

Um número diferente de amostras foi coletado para cada área teste, por dependerem da existência das feições na região e da possibilidade de obter amostras espectro-temporais

consistentes. Assim, para a área Sul foram coletados 1.403 pontos, para 18 classes; para a área Central foram 1.363 pontos, para 16 classes; e para a área Norte foram 397 pontos envolvendo 30 classes. A área Norte da MA é um desafio para esta abordagem, uma vez que a ocorrência de nuvens durante todo o ano [15], implica em um número menor de áreas possíveis de serem amostradas.

A coleta de amostras com base na análise em diferentes períodos do ano permite amostras mais detalhadas e robustas [19]. Os resultados de acurácia geral e kappa (k-fold) foram respetivamente: 0,79 e 0,77 na área Norte, 0,95 e 0,94 nas áreas Central e Sul.

Os índices e bandas utilizados, inicialmente escolhidos com base nas bandas *defaut* do pacote SITS e refinados em um processo de iteração contínua, apresentaram desempenhos consistentes nas três áreas. Entre eles, SAVI, NBR, NDVI, NDWI, B02, B04, B08, B09, B11 e B12 resultaram em gráficos espectro-temporais que evidenciaram maior distinção entre classes, sugerindo uma separabilidade aprimorada entre feições no mapa [12].

O mapa da área Sul foi gerado com a adição do índice MSI (*Moisture Stress Index*). Por ser um índice que mede o estresse hídrico [20], ele tenha sido útil em áreas com predominância de fitofisionomias de estepe gramíneo-lenhosa [13]. Na área Norte, além do MSI, houve a adição do índice NDRE (*Normalized Difference Red Edge*) que analisa o vigor da vegetação através do conteúdo da clorofila, sendo útil na separabilidade das vegetações de ecótono entre os biomas Mata Atlântica e Caatinga [13].

As iterações amostrais melhoraram com a adição de cada índice e banda, e pioraram com sua remoção. Além disso, observou-se que as combinações influenciam o comportamento espectral, sendo que um índice apresenta resposta diferente quando analisado isoladamente ou em conjunto [21].

No geral os modelos representaram bem comportamentos de mudança, como supressão em áreas de floresta, por exemplo. Porém, gerou resultados que precisam ser adequados ao projeto, como regiões de silvicultura com episódios de corte que foram classificadas como incrementos de desmatamento em 2023. Contudo, para o caso do emprego da metodologia no sistema Prodes, as classes de silvicultura provavelmente estariam sob a "máscara" de desmatamento [9], e sua representação para a análise final seria mínima.

Apesar dos valores de acurácia das amostras na região Norte serem mais baixos, observou-se *clusters* com boa representatividade de classes de desmatamentos trimestrais, área natural, floresta, hidrografia e cultivo.

As classes afloramento rochoso, área urbana e desmatamento_2022, apresentaram confusões nas áreas Norte e Sul. Isto se deve a comportamentos espectro-temporais semelhantes da rocha exposta, que perante a incidência solar reflete padrão similar às coberturas de concreto, bem como o solo exposto arenoso [22]. Houve confusões entre as classes de desmatamento trimestral e cultivo na região Sul. Nestes locais, o cultivo apresenta vegetação baixa no início do ano, confundido com áreas campestres [13], e, na colheita, o solo exposto é interpretado como supressão.

Na região Central, houve confusões entre classes de floresta, áreas naturais, pastagem e desmatamento_2022 devido à predominância de florestas estacionais [13]. As mudanças espectro-temporal em períodos de seca em florestas que perdem folhas (estacionais) se assemelham às pastagens, algumas áreas desmatadas e áreas não florestais. A subdivisão da classe florestal está em teste, sugerindo-se ainda agregar à metodologia o emprego de limites com base nas fitofisionomia ou ecorregiões. Contudo, a acurácia da classe floresta (k-fold) de usuário chegou 96% e do produtor a 99%.

Apesar de indesejadas, imagens com áreas com nuvens são muito comuns na MA [14]. Não por acaso as nuvens influenciam muito na coleta de amostra e análise dos *clusters* bem como no mapa de probabilidade [11]. Os mosaicos usam critérios de escolha de melhor pixel com banda SCL, que distingue nuvem e sombra de vegetação e água, porém a ocorrência de nuvem e sombra ainda acontece em grandes proporções. Estes alvos distorcem a resposta espectral das feições, podendo levar áreas naturais a serem classificadas como supressão da vegetação nativa.

Diante destas dificuldades, coletar amostras do considerando o comportamento temporal de nuvem para representar a paisagem apresentou resultados satisfatória. Desse modo, a coleta de amostra de floresta e floresta com ocorrência de nuvens se deu com a classe de floresta sendo composta por amostras nas imagens que apresentem feição de floresta na primeira e na última imagem, tendo ou não ocorrência de nuvens no restante das imagens trimestrais; e a classe floresta_nuvem, composta por amostras em regiões que havia predominância de feições de floresta, porém com ocorrência de nuvem na primeira e/ou na última imagem do mosaico trimestral.

3.2. Acurácia do mapa de probabilidade

As acurácias gerais dos mapas foram de 76% na área Norte, 87% na Central e 86% na Sul. Na acurácia ponderada por classe, nas classes de desmatamentos trimestrais, os erros de omissão superaram os de comissão, o que foi confirmado pela análise visual dos mapas (Figura 3). Em contrapartida as demais classes incluídas áreas previamente desmatadas, os erros por comissão são maiores para as três áreas.

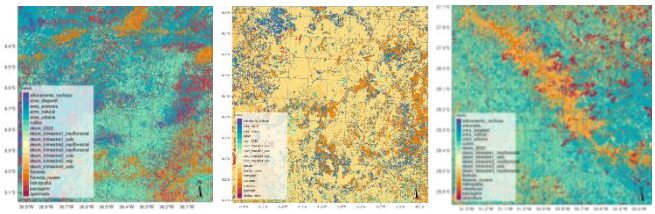


Figura 3. Resultado dos mapas de probabilidade das imagens classificadas.

A partir da análise das áreas, percebeu-se a necessidade de uma revisão do conceito da coleta de amostra das classes de cultivo, desmatamento_2022 e desmatamento com solo exposto no primeiro trimestre de 2023. As três classes apresentam respostas espectrais parecidas para a primeira imagem mosaico utilizada (solo exposto), o que pode justificar a baixa acurácia do produtor do mapa de probabilidade (0,04 a 0,40) para as 3 classes mencionadas.

Auditorias visuais, bem como estatísticas comparativas entre os resultados dos modelos e o PRODES-MA de 2023 são previstos a fim de entender quais mudanças serão implementadas. Estas análises serão executadas assim que os testes aqui descritos atingirem acurácia igual ou superior a empregada na metodologia atual.

5. CONCLUSÕES

A abordagem semiautomática aplicada à MA utilizando o pacote SITS processou séries temporais do Sentinel-2, com imagens de 10 m de resolução, cobrindo 11.151,36 km² em três regiões. Foram coletados 3.163 pontos amostrais, com acurácia geral entre 79% e 95% e resultados de Kappa acima de 0,77 e acurácias gerais do mapa de probabilidade de 76% (Norte), 87% (Central) e 86% (Sul). Os mapas gerados identificaram boa separabilidade em áreas naturais, em mudanças provocadas pela supressão, que permite a possível identificação de desmatamento trimestral. Ainda que a ocorrência de nuvens seja uma questão a ser trabalhada na metodologia, e que existam classes que resultem em confusões pois possuem respostas espectro-temporais similares, os resultados destacam a viabilidade de aplicação em larga escala para aprimorar o monitoramento ambiental. Sendo esse um processo de iteração contínua, percebe-se resultados promissores para o emprego do mapeamento semiautomático da supressão da vegetação nativa da MA.

5. AGRADECIMENTOS

Agradecimento ao CNPq processo 422354/2023-6 (Monitoramento e avisos de mudanças de uso e cobertura da terra nos biomas brasileiros – Capacitação e semiautomatização do Programa BiomasBR), sediado e executado pelo Instituto Nacional de Pesquisas Espaciais (INPE).

8. REFERÊNCIAS

[1] R. A. Mittermeier, et al. Fonseca. Hotspots revisited: earth's biologically richest and most endangered terrestrial ecoregions, 2nd edn. *CEMEX*, Mexico City, 2005.

[2] K. J. Williams, et al. Forests of East Australia: the 35th biodiversity hotspot. In: Zachos FE, Habel JC (eds) *Biodiversity hotspots*. Springer, Heidelberg, p 295–310, 2011

[3] F. R. SCARANO, P. CEOTTO. Brazilian Atlantic Forest: impact, vulnerability, and adaptation to climate change. *Biodiversity and Conservation*, v. 24: n. 9, p. 2319-2331, 2015. DOI: 10.1007/s10531-015-0972-y

[4] C. Béllard, et al. Vulnerability of biodiversity hotspots to global change. *Global Ecology Biogeography*, v. 23: p. 1376–1386, 2014. DOI: 10.1111/geb.12228.

[5] Instituto Nacional de Pesquisas Espaciais. (n.d.). TerraBrasilis. Disponível em: https://terrabrasilis.dpi.inpe.br/

[6] J. Romanelli, et al. Biodiversity responses to restoration across the Brazilian Atlantic Forest. *Science of The Total Environment*, v. 821: p. 153403, 2022.

[7] S. Amaral, M. M. dos S. Cursino; C. A. de Almeida. Monitoring Atlantic Forest Deforestation By Remote Sensing Systems. In: *Anais Do XX Simpósio Brasileiro De Sensoriamento Remoto*, v.20: p.464 – 467, 2023.

[8] F. O. Passos, et al. Efeito Da Resolução Espacial Sobre a Série Histórica De Monitoramento De Desmatamento Da Mata Atlântica. In: *Anais Do XX Simpósio Brasileiro De Sensoriamento Remoto*, v.20: p.1598 – 1601, 2023.

[9] R. Z. Molinez, et al. Historical and recent deforestation patterns in the Atlantic Forest from satellite monitoring system – PRODES. *Revista Brasileira de Cartografia*, v. 76, 2024. DOI: 10.14393/rbcv76n0a-72589.

[10] [14] IBGE. Contas de Ecossistemas: Uso da Terra nos Biomas Brasileiros (2000-2018), IBGE, Rio de Janeiro, 2020. 95p, 2020

[11] F.O. Passos, et al. PRODES Mata Atlântica: discussing the digital transition from visual interpretation to semi-altomatic detection of forest removal. *XXIV Brazilian Symposium on Geoinformatics*,: p. 298 – 309, 2023.

[12] R.Simoes, Camara, G. et al. Satellite Image Time Series Analysis for Big Earth Observation Data. *Remote Sensing*, v.12, n. 17: p2735.

[13] IBGE. Manual técnico da vegetação brasileira (2012), IBGE, Rio de Janeiro. 95p, 2020

[14] BDC Explorer. Disponível: https://data.inpe.br/bdc/explorer/explore.

[15] INPE - FUNCATE, Almeida et al. Monitoramento Ambiental dos Biomas Brasileiros por Satélites: Mata Atlântica, Caatinga, Pampa e Pantanal - Relatório de Referência Metodológica dos Subprojetos 1 a 4 (v1), 2022. Disponível em: http://biomas.funcate.org.br/.

[16] F. Petitjean, J. Inglada, and P. Gancarski, "Satellite Image Time Series Analysis Under Time Warping. *IEEE Transactions on Geoscience and Remote Sensing*, v. 50, n. 8: p. 3081–3095, 2012. Doi: 10.1109/TGRS.2011.2179050.

[17] W. G. Cochran. Sampling Techniques. Copyright, USA, 1977. Disponívelem: https://fsapps.nwcg.gov/gtac/CourseDownloads/IP/Cambodia/FlashDrive/Supporting_Documentation/Cochran_1977_Sampling%20Techniques.pdf.

[18] D. Olofsson, et al. Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, v. 51, n. 11: p. 933 – 938. DOI: 10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2.

[19] Z. Gao, et al. Training sample selection for robust multi-year within-season crop classification using machine learning. *Computers and Electronics in Agriculture*, v. 214: p. 107927, 2023. DOI: 10.1016/j.compag.2023.107927.

[20] R. Coluzzi, et al. A first assessment of the Sentinel-2 Level 1-C could mask producto to support informed surface analyses. *Remote Sensing and Environment*, v. 217: p. 426-443, 2018. DOI: 10.1016/j.rse.2018.08.009.

[21] R. C. F. Silva, et al. Análise Comparativa De Índices Espectrais Para a Extração Automática da Lâmina d'água em Imagem do Satélite Landsat8OLI. In: *Anais Do XIX Simpósio Brasileiro de Sensoriamento Remoto*, v.19: p. 96629, 2019.

[22] B. Rivard, et al. Varnish, sediment, and rock controls spectral reflectance of outcrops in arid regions. *Geology*, v.20, n.4: p. 295-298, 1992. DOI:10.1130/0091-7613(1992)020<0295:VSARCO>2.3.CO;2.