

APRENDIZADO DE MÁQUINA E USO DE SÉRIES TEMPORAIS DE IMAGENS SENTINEL-2 NO MONITORAMENTO DA COBERTURA E USO DA TERRA

André Silva Tavares¹, Filipe Castro Felix¹, Felipe Souza², Alexandre Camargo Coutinho³, João Francisco Gonçalves Antunes³, Júlio César Dalla Mora Esquerdo³

¹Bolsista CNPq Projeto TerraClass Amazônia, andre.tavares@colaborador.embrapa.br, filipe.felix@colaborador.embrapa.br, ²Instituto Nacional de Pesquisas Espaciais, Av. dos Astronautas, 1758, CEP 12227-010, São José dos Campos, SP, Brasil, felipe.carvalho@inpe.br, ³Embrapa Agricultura Digital, Av. André Tosello, 209, CEP 13083-886, Campinas, SP, Brasil, alex.coutinho@embrapa.br, joao.antunes@embrapa.br, julio.esquerdo@embrapa.br

RESUMO

O uso de séries temporais de imagens de satélite aliadas a algoritmos de aprendizado de máquina é uma abordagem que vem sendo aprimorada nos últimos anos. Considerando a acelerada dinâmica de expansão da fronteira agropecuária em regiões da Amazônia brasileira, este estudo teve como objetivo identificar alterações no uso da terra na região noroeste do Maranhão e leste do Pará, por meio de séries temporais do Sentinel-2 e o algoritmo *Random Forest*. Foram coletadas amostras aleatórias do mapa TerraClass 2014 e, para aprimorar o conjunto amostral, utilizou-se o algoritmo de mapas auto-organizáveis. Os resultados mostraram que as transições mais expressivas, entre 2014 e 2020, ocorreram em áreas de agricultura temporária, com aumento de 265%, e silvicultura, com 176%. O protocolo metodológico promoveu ganhos importantes frente à performance dos mapeamentos convencionais e mostrou ter potencial para auxiliar no monitoramento das mudanças no uso da terra no território nacional.

Palavras-chave — Amazônia Legal, *Random Forest*, Sensoriamento Remoto.

ABSTRACT

The use of time series of satellite images combined with machine learning algorithms is an approach that has been improved in recent years. Considering the accelerated dynamics of expansion of the agricultural frontier in regions of the Brazilian Amazon, this study aimed to identify changes in land use in the northwest region of Maranhão and eastern Pará, using Sentinel-2 time series and the Random Forest algorithm. Random samples were collected from the TerraClass 2014 map and, to improve the sample set, the algorithm of self-organizing maps was used. The results showed that the most significant transitions, between 2014 and 2020, occurred in areas of temporary agriculture, with an increase of 265%, and forestry, with 176%. The methodological protocol promoted important gains compared to the performance of conventional mappings and showed the potential to assist in the monitoring of changes in land use in the national territory.

Key words — Legal Amazon, *Random Forest*, Remote Sensing.

1. INTRODUÇÃO

O Brasil é um dos maiores exportadores de commodities agropecuárias do mundo. Contudo, um dos desafios da produção de alimentos para atendimento da demanda mundial é reduzir suas taxas de desmatamento e emissões de gases de efeito estufa, equalizando ganhos econômicos e práticas agrícolas sustentáveis [1]. Nesse contexto, o monitoramento do uso e cobertura da terra torna-se uma ferramenta estratégica para apoiar o planejamento e a gestão territorial.

Na detecção de mudanças de uso da terra, métodos de processamento de imagens, sobretudo aquelas disponibilizadas pelo sensoriamento remoto, desempenham um papel determinante na identificação, mapeamento, avaliação e monitoramento [2]. Avanços nas metodologias para classificação automática a partir de séries temporais de imagens de satélite têm contribuído para a melhoria na acurácia dos modelos [2,1]. A construção de ambientes dotados de alta dimensão espaço-temporal que permitam utilizar de forma ampla os conjuntos de valores de uma determinada série temporal, caracterizam uma abordagem mais robusta que vem sendo aprimorada, para operar com grandes volumes de dados, mais notadamente com os avanços do aprendizado de máquina [1].

O desenvolvimento do aprendizado de máquina é uma realidade presente em vários ramos da ciência. Em sensoriamento remoto os modelos são capazes de identificar e distinguir perfis temporais de índices ou bandas espectrais, melhorando a exatidão na classificação conforme a robustez das amostras. No entanto, alguns desafios devem ser levados em consideração, tais como: características físicas específicas do bioma, sazonalidade climática, cobertura de nuvens, transição entre diferentes fitofisionomias, entre outros [3].

A região da Amazônia Legal Brasileira possui um território com aproximadamente 5 milhões de km², representando quase 59% do território brasileiro. Entre as porções noroeste do Maranhão e leste do Pará situa-se uma das regiões de maiores índices de expansão da fronteira agrícola, estimulada pela adoção de tecnologias que proporcionam alta produtividade e com investimento massivo do mercado fundiário pelo capital financeiro [7,8].

O leste paraense pode ser compreendido como uma continuidade da realidade geográfica consolidada pela região do Matopiba (acrônimo que agrupa Maranhão, Tocantins, Piauí e Bahia). Nessa região, o avanço agrícola acelerou a partir da década de 2000, quando já atingia um total de área plantada com soja de 100 mil hectares, chegando a 1,4 milhões de hectares em 2016 [4,5].

Nesse contexto, este estudo teve como objetivo desenvolver um modelo de aprendizado de máquina para classificação de cobertura e uso da terra a partir de séries temporais de imagens Sentinel-2. Esse modelo visou aprimorar a classificação e identificar áreas de mudanças na cobertura e uso da terra entre as regiões noroeste do Maranhão e leste do Pará, entre 2014 e 2020.

As áreas mapeadas no ano base de 2020 foram comparadas com aquelas já existentes no ano de 2014 e os resultados evidenciaram a ocorrência de um aumento expressivo das áreas de agricultura temporária, silvicultura e das áreas de pastagem na região. Os resultados foram representativos e satisfatórios, proporcionando um bom nível de detalhamento das classes de interesse, sendo viável para o monitoramento e detecção de mudanças no uso da terra.

2. MATERIAL E MÉTODOS

Séries temporais de imagens de satélite e aprendizado de máquina para a detecção de mudanças de uso da terra é uma abordagem em desenvolvimento na última década [4,5,6]. No Brasil, o projeto BDC (*Brazil Data Cube*), do Instituto Nacional de Pesquisas Espaciais (INPE), disponibiliza um banco de dados que inclui imagens CBERS-4, Landsat-8, Sentinel-2 e MODIS (Terra e Aqua).

Equipes do INPE também desenvolveram o pacote computacional SITS (*Satellite Image Time Series Analysis for Earth Observation Data Cubes*). Escrito em linguagem R, o pacote possui código aberto e possibilita a análise, visualização e classificação de dados de séries temporais, a partir de coleções de imagens disponíveis em serviços de nuvem e aprendizado de máquina [6].

2.1 Pacote SITS

O SITS faz uso de coleções de imagens organizadas em cubos de dados, onde cada localização espacial está associada a uma série temporal [6]. Locais com rótulos conhecidos treinam um classificador de aprendizado de máquina, que classifica a série temporal do cubo de dados. Neste estudo foram utilizadas séries temporais de imagens Sentinel-2 para o ano safra 2019/2020 (Tabela 1).

Em seguida foi empregado o algoritmo *Random Forest* para o treinamento e a geração do mapa de uso da terra. Considerando o conjunto amostral selecionado, a Figura 1 destaca a importância dos atributos temporais na técnica de *Random Forest*.

Base	Dados
Fonte	Brazil Data Cube
Coleção	S2-SEN2COR 10 16D STK-1
Bandas	EVI, NDVI, B02, B03, B04, B05, B06, B07, B08, B8A, B11, B12, CLOUD
Blocos	087087, 087088, 087089, 088087, 088088, 088089
Início	28-07-2019
Fim	16-11-2020

Tabela 1. Base de informações usadas para elaboração do cubo de dados Sentinel-2.

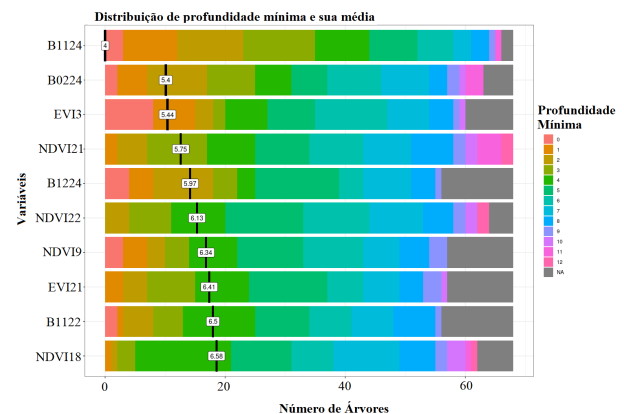


Figura 1. Distribuição de profundidade mínima das árvores e variáveis no treinamento do modelo Random Forest.

A ordenação das variáveis (eixo Y) é feita com base na média de profundidade mínima de cada árvore. A profundidade mínima de cada árvore é a quantidade de nós que essa árvore possui em nível de altura. Logo, as árvores com profundidade zero escolhem os atributos destacados como raiz. Por exemplo, o algoritmo *Random Forest* selecionou a variável B11 (SWIR 1) como a mais representativa no tempo 24. Nesse caso, a média de profundidade mínima de árvores em B1124 (banda/tempo) é de quatro árvores.

2.2 Controle de Qualidade das Amostras

Nessa etapa, é necessário o uso de métodos de pré-processamento para melhorar a qualidade das amostras, removendo rótulos incorretos ou de baixo poder discriminatório [6]. O SITS incorporou o algoritmo de agrupamento baseado em Mapas Auto-Organizáveis (SOM), disponibilizando um protocolo de redução de dimensionalidade, onde dados de alta dimensão espectro-temporal são mapeados em um mapa bidimensional, mantendo as relações topológicas entre os padrões de dados [6].

O SOM é constituído por neurônios com dimensão igual às amostras que serão usadas no treinamento do modelo. Com um valor aleatório inicial, os neurônios são posteriormente treinados por aprendizado competitivo. O algoritmo calcula as distâncias de cada membro do conjunto de treinamento para os neurônios e encontra o neurônio

mais próximo da entrada, chamado de unidade de melhor correspondência [6].

Neste estudo foram extraídas amostras aleatórias das classes obtidas pelo projeto TerraClass Amazônia (2014), são elas: *agricultura semiperene*, *agricultura temporária*, *pastagem arbustiva*, *pastagem herbácea*, *silvicultura*, *vegetação florestal natural e água*. A classe de agricultura temporária foi subdividida em duas classes: agricultura temporária de 1 ciclo e agricultura temporária de 2 ou mais ciclos, usando a resposta espectral das amostras analisadas no SOM.

No total, foram agrupadas 7.215 amostras pelo algoritmo SOM, a partir da variação da resposta espectral (Figura 2) sob vinte e cinco cenas (2019 a 2020), dez bandas Sentinel-2 e dois índices (NDVI e EVI), descritas na Tabela 1, que resultaram em trezentas dimensões espectro-temporais.

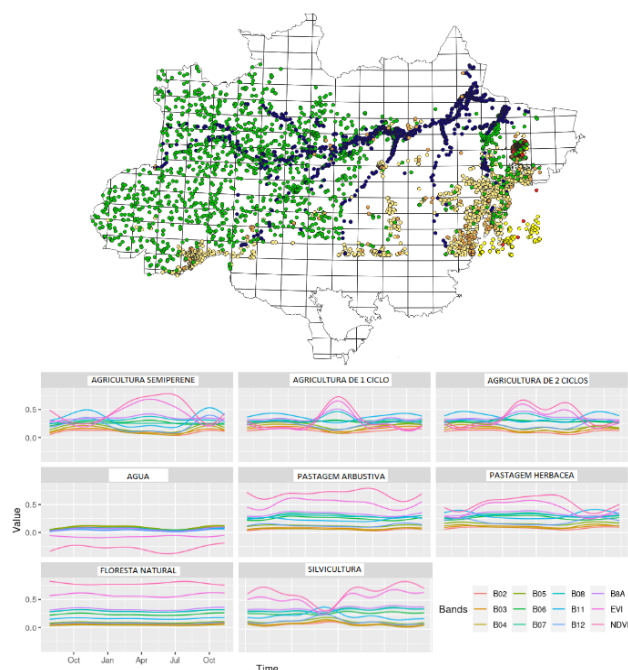


Figura 2. Distribuição e perfil espectro-temporal das amostras de treinamento para o modelo *Random Forest*.

Nas áreas de silvicultura, devido ao seu ciclo de corte ser mais longo em relação às outras culturas, o cubo de dados do ano safra 2019/2020 não foi suficiente para sua discriminação, principalmente na separação entre silvicultura e áreas de florestadas natural. Logo, foi considerado o período entre 2017/2020, ou setenta e cinco datas (cenas), que resultou em novecentas dimensões espectro-temporais. Os neurônios formaram um conjunto de *clusters*, preservados pela variável topológica de vizinhança.

As informações sobre a variação intraclases e interclasses foram usadas para detectar amostras ruidosas (*outlier*). Assim, as amostras que não correspondiam às classes foram removidas antes do treinamento (Figura 3).

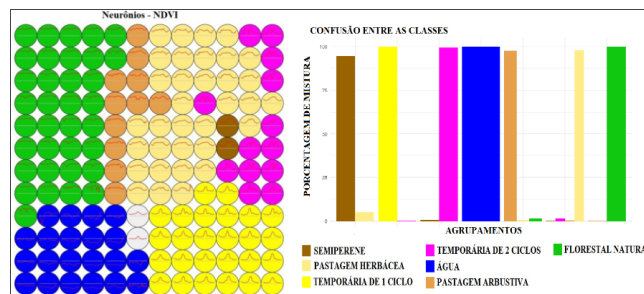


Figura 3. Neurônios e seus *clusters* e as confusões entre as classes de interesse.

3. RESULTADOS E DISCUSSÃO

A mistura nos neurônios alcançou, no máximo, 4,5%, referente às classes de cultura semiperene e pastagem herbácea, o que já era esperado, visto que o perfil espectro-temporal das duas classes é muito semelhante. As classes restantes não ultrapassaram 2% de confusão entre si. A classificação resultante é observada na Figura 4.

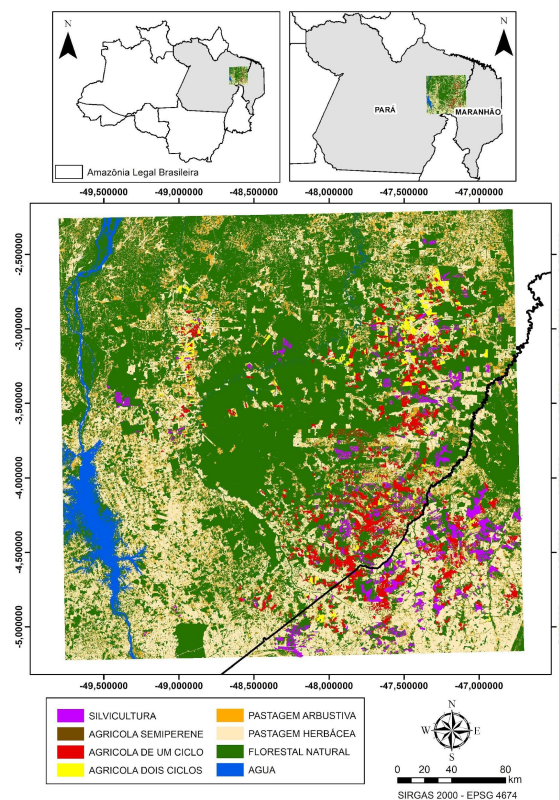


Figura 4. Classificação de uso da terra na região noroeste do Maranhão e leste do Pará entre 2019 e 2020.

A classe de agricultura temporária apresentou o aumento de área mais expressivo entre os anos de 2014 e 2020, correspondendo a 265%, seguida pela silvicultura, com 176%. A agricultura semiperene (cana-de-açúcar) permaneceu pouco alterada no período, com acréscimo de em 1% da sua área (Figura 5).

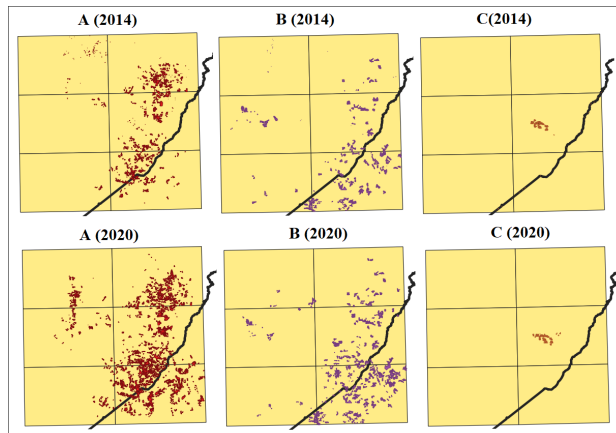


Figura 5. Agricultura temporária (A), silvicultura (B) e agricultura semiperene (C) nos anos de 2014 e 2020.

A expansão de áreas agrícolas e silvicultura se dá na transição de áreas de pastagem e desflorestamento. A pastagem herbácea é a classe de maior expressão territorial na região, cobrindo 44,84% e 37,97%, nos anos de 2014 e 2020, respectivamente (Tabela 2).

Uso da terra	Área (M/ha)	Área (%)	Período
Agrícola Semiperene	0,016	0,22	2014
	0,016	0,22	2020
Agrícola Temporária	0,198	3,28	2014
	0,525	7,36	2020
Silvicultura	0,140	2,32	2014
	0,247	3,46	2020
Pastagem Herbácea	2,699	44,84	2014
	2,708	37,97	2020
Pastagem Arbórea	1,170	19,43	2014
	1,334	18,71	2020
Vegetação Secundária	1,793	29,78	2014
	2,301	32,26	2020

Tabela 2. Uso da terra entre o período de 2014 a 2020.

Com base no sensor MODIS (250 m) e interpretação visual, o mapeamento do projeto TerraClass 2014 contou com um nível de detalhe menor, quando comparado a este estudo. Foi verificado que ocorreram omissões, por exemplo, na classe de vegetação secundária (VS) em matas de galeria lineares, no entorno de rios de menor largura. Sendo assim, essa classe temática acabou sendo subestimada.

Com o sensor MSI (10 m) do Sentinel-2, o acréscimo de VS era esperado no mapa de 2020. Todavia, as áreas de VS de 2014 que não foram detectadas no mapa de 2020, foram na ordem de 0,54 M ha. Isto é, áreas que foram desflorestadas, dando origem a áreas de cultivo diversas. Por outro lado, ocorreu a inclusão de VS no mapa de 2020, na ordem de 1,04 M ha, devido à maior resolução espacial.

4. CONCLUSÃO

Deve-se destacar que o uso de séries temporais de imagens Sentinel-2 e o algoritmo *Random Forest* permitiu detectar mudanças no uso da terra com maior representatividade, quando comparados a mapas anteriores. Logo, foi possível observar que o monitoramento do uso da terra na fronteira dos Estados do Maranhão e Pará é de especial importância, visto que os resultados demonstraram, por exemplo, que áreas agrícolas temporárias quase triplicaram, entre 2014 a 2020, e ao mesmo tempo, áreas de vegetação secundária foram reduzidas.

AGRADECIMENTOS

Ao Centro Gestor e Operacional do Sistema de Proteção da Amazônia (Censipam) (Projeto TerraClass Amazônia/TED Censipam-CNPTIA); ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); ao projeto *Brazil Data Cube* (INPE); e aos desenvolvedores do pacote computacional SITS.

5. REFERÊNCIAS

- [1] M. Picoli, G. Camara, L. Sanches, R. Simões, A. Carvalho, A. Maciel, A. Coutinho, J. Esquerdo, J. Antunes, R. Begotti, D. Arvor, C. Almeida. Big earth observation time series analysis for monitoring Brazilian agriculture. *ISPRS journal of photogrammetry and remote sensing*. v. 145, pp. 328-339, 2018.
- [2] L. Santos, K. Ferreira, G. Camara, M. Picole, R. Simoes. Quality control and class noise reduction of satellite image time series. *Journal of Photogrammetry and Remote Sensing*. v. 177, pp. 75-88, 2021.
- [3] D. Cho, S. Schwaida, R. Cicerelli, T. Almeida, A. Ramos, E. Sano. Desempenho do Algoritmo de Classificação de Imagens Random Forest para Mapeamento do Uso e Cobertura do Solo no Cerrado Brasileiro. *Anuário do Instituto de Geociências*, v. 44, pp. 1-11, 2021.
- [4] S. Frederico, M. Almeida. Capital financeiro, land grabbing e multiescalaridade na grilagem de terra na região do MATOPIBA. *Revista NERA*. v. 22, pp. 123-147, 2019.
- [5] IBGE, Instituto Brasileiro de Geografia e Estatística. *Produção Agrícola Municipal*, 2018. Disponível em: <https://sidra.ibge.gov.br/tabela/1612>. Acesso em: 20 de outubro de 2022.
- [6] R. Simões, G. Camara, G. Queiroz, F. Souza, P. R. Andrade, L. Santos, A. Carvalho, K. Ferreira. Satellite Image Time Series Analysis for Big Earth Observation Data. *Remote Sens*. v. 13, pp. 2428, 2021.
- [7] R. Freitas. Brazil's agricultural expansion Main crops in Matopiba. *Revista de Política Agrícola*, v. 1, pp. 70-85, 2022.
- [8] M. Araújo, E. Sano, E. Bolfe, J. Santos, J. Santos, F. Silva. Spatiotemporal dynamics of soybean crop in the Matopiba region, Brazil (1990–2015). *Land Use Policy*, v. 80, pp. 57-67, 2019.