

CLASSIFICAÇÃO DE USO E COBERTURA DA TERRA EM ÁREAS DE NÃO FLORESTA DO SUDESTE PARAENSE ATRAVÉS DA PLATAFORMA GOOGLE EARTH ENGINE (GEE)

Larisse Fernanda Pereira de Souza¹, Marcos Adami², Jones Remo Barbosa Vale³, Igor dos Santos e Silva⁴, Luiz Cortinhas Ferreira Neto⁵, Ingrid Cássia Lima Porto,⁶Alessandra RodriguesGomes⁷

¹Universidade Federal do Pará, Rua Augusto Corrêa, N°1, larisse.souza14@gmail.com;²Instituto Nacional de Pesquisas Espaciais/ Centro Regional da Amazônia, Av. Presidente Tancredo Neves, N° 2651, marcos.adami@inpe.br. ³Universidade Federal Rural da Amazônia, Av. Presidente Tancredo Neves, N° 2501, jonesremo@hotmail.com;⁴Universidade Federal Rural da Amazônia, Av. Presidente Tancredo Neves, N° 2501, igorssilva20@gmail.com ;⁵Universidade Federal do Pará, Rua Augusto Corrêa, N°1, Luizcf14@gmail.com⁶Universidade Federal do Pará, Rua Augusto Corrêa, N°1, Ingrid.cassia@gmail.com⁷Instituto Nacional de Pesquisas Espaciais/ Centro Regional da Amazônia, Av. Presidente Tancredo Neves, N° 2651, alessandra.gomes@inpe.br;

RESUMO

Este trabalho tem como objetivo apresentar um método de mapeamento da formação NF localizada no sudeste paraense a partir da base de dados da plataforma Google Earth Engine (GEE). Para a realização do mapeamento foram utilizadas imagens do satélite Landsat 5, 7, 8 e do MODIS para os anos de 2001, 2010 e 2017. O mapeamento foi realizado através da classificação supervisionada com o algoritmo *Random Forest*, obteve-se uma precisão global de 68%, no qual a classe a classe Savana Arborizada teve um alto índice de omissão, correspondendo a 0,94. A classe de pastagem foi confundida com a classe de savana apresentando o maior erro de inclusão. Portanto, o GEE através do algoritmo *Random Forest* mostrou-se uma ferramenta eficaz no mapeamento e classificação do uso da terra em áreas de NF.

Palavras-chave — *Classificação supervisionada Random Forest, Google Earth Engine*

ABSTRACT

The objective of this paper is to present a method of mapping NF formation located in southeastern Pará from the database of the Google Earth Engine (GEE) platform. In order to perform the mapping, the Landsat 5, 7, 8 and MODIS satellite images were used for the years 2001, 2010 and 2017. The mapping was performed through the supervised classification with the Random Forest algorithm, obtaining an overall accuracy of 68%, in which the wooded savanna class had a high omission rate, corresponding to 0.94. The pasture class was confused with the savannah class presenting the largest inclusion error. Therefore, GEE using the Random Forest algorithm has proved to be an effective tool in the mapping and classification of land use in NF areas.

Key words — *Supervised classification, Random Forest, Google Earth Engine*

1. INTRODUÇÃO

A Amazônia é o maior ecossistema florestal tropical contínuo na Terra e sem dúvida um ícone global, com extensão total de aproximadamente 5,4 milhões km² e presente em 9 países do continente sul-americano, sendo que mais de 62% da área deste ecossistema encontra-se em território brasileiro. Abriga 25% de todas as espécies terrestres no mundo, e responsável por 15% de toda fotossíntese terrestre global [1]. Contudo, apesar de ser composto por grandes extensões florestais é ingênuo acreditar que este ecossistema é homogêneo e uniforme. Na verdade, este é um ambiente complexo com formações de diversos ecótonos, o que gera grande diversidade florística [2].

Neste complexo, as formações vegetais não florestais recobrem uma área de 290 mil km² do Bioma Amazônia brasileiro, aproximadamente 7% da sua área e, é composta principalmente por campinaranas e savanas [3].

Estudos de mapeamento das fitofisionomias Floresta/Não Floresta são muito importantes para a utilização dos recursos naturais. E isto, planejamento e gestão eficaz dos recursos naturais vem se ampliando consideravelmente a partir de técnicas de sensoriamento remoto [4].

As áreas florestais da Amazônia são monitoradas desde 1988 quando foi criado o Projeto de Monitoramento e Desmatamento da Amazônia (PRODES) para obter taxas anuais de desflorestamento bruto da Amazônia Legal Brasileira fornecendo informações detalhadas sobre o desmatamento. Porém o projeto PRODES não monitora as áreas de formação Não Florestais [5], o que restringe informações sobre essas áreas e sua diversidade ambiental e grau de antropização.

As savanas amazônicas são de extrema relevância para a conservação da biodiversidade, sendo compostas por comunidades vegetacionais de numerosas espécies endêmicas. No entanto, as savanas amazônicas são pouco estudadas, onde têm sido isoladas pela própria floresta, e apesar dos estudos serem restritos, eles mostram claramente a heterogeneidade existente na composição das comunidades de flora e fauna dessas áreas [6].

Desta maneira, o objetivo deste trabalho é apresentar um método de mapeamento de uma área de formação não florestal localizada no sudeste paraense, utilizando a plataforma Google Earth Engine (GEE).

2. MATERIAIS E MÉTODOS

2.1 Área de Estudo

Para realização deste trabalho foi selecionada como área de análise uma região de formação Não Florestal localizada nos municípios de Rio Maria, Redenção, Floresta do Araguaia, Conceição do Araguaia, Santa Maria das Barreiras, Pau D'arco e Santana do Araguaia, região sudeste do Estado do Pará, Amazônia, que abrange uma área de 7.710 Km². A área possui clima tropical chuvoso, quente e úmido. Sua cobertura vegetal compõe-se principalmente, por uma tipologia de transição Floresta/Cerrado [7].

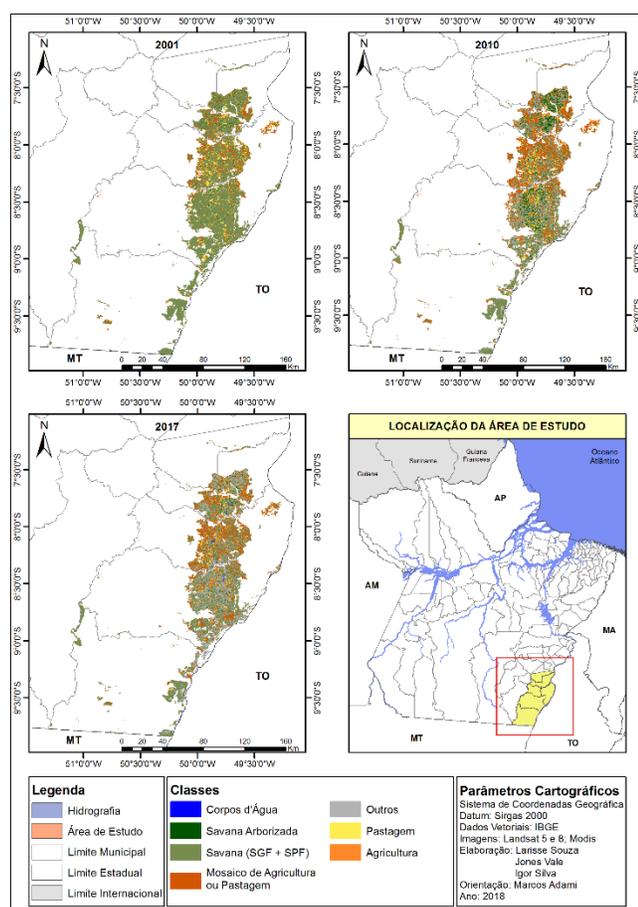


Figura 1. Localização e classificação do uso e cobertura da terra para os anos de 2001, 2010 e 2017

2.2 Obtenção de Dados

Neste trabalho, foram utilizadas imagens do satélite Landsat 5,7 e 8 que apresentaram melhores resultados da

geração dos índices de vegetação utilizados, sendo estes: NDVI, NDWI e WVI ou NDDI, VrNIR red.

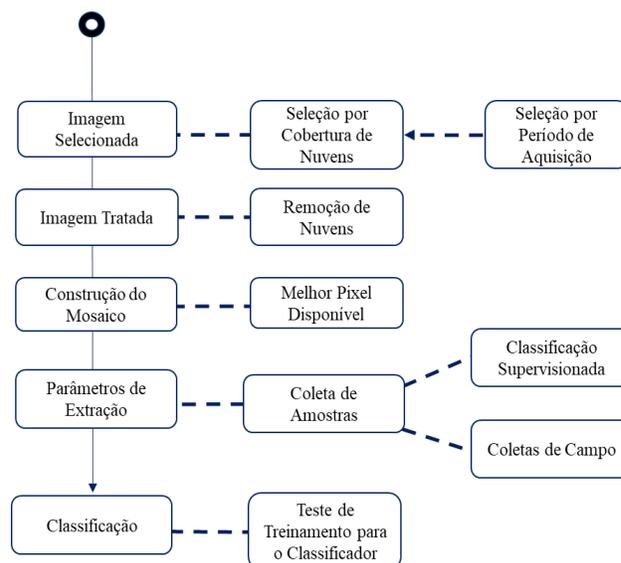


Figura 2. Fluxograma de etapas para execução da metodologia

A classificação supervisionada por meio do algoritmo Random Forest foi realizada sobre áreas de controle, estas foram escolhidas por possuírem grande diversidade ecológica, isto objetiva obter nas amostras uma maior representatividade de toda diversidade no bioma Amazônia no que se refere a áreas com cobertura vegetal não florestal, por isso as etapas pós construção:

1. *Preparação dos Mosaicos*: Os mosaicos anuais consistem em medianas trimestrais de composição sem nuvem, para a composição de mosaico sem nuvem foi utilizado o algoritmo Landsat Simple Cloud Score, e para diminuição de névoa e neblina foi utilizado o *Best Available Pixel* (BAP) [8].
2. *Coleta de amostras*: Visando melhor acurácia de algoritmo supervisionado, se fez necessário um conjunto de amostras obtidas por especialista.
3. *Classificação*: Etapa de uso do algoritmo em todo mosaico, já treinado com 60% do conjunto de amostras.
4. *Teste de Treinamento para o Classificador*: 40% do conjunto de amostras, sendo estes não contidos no conjunto de treinamento, foram utilizadas para compor o conjunto de validação.

A primeira fonte de classificação “Landsat + Random Forest” classifica a cobertura do solo usando ferramentas de pré-processamento de amostras compostas por SMOTE e Threshold Limits para engajar o equilíbrio da classe, afinal extraímos 5 classes para representar a diversidade das áreas NF: Agricultura, Outros, Pasto, Savanas (Sgf + Spf), as nomenclaturas Sgf e Spf estão classificadas de acordo com o IBGE como Savana Gramíneo Lenhosa (Sgf) e Savana Parque (Spf) ambas com floresta de galeria [3]. A última classe apresentada é Savana Arborizada que foi classificada

pelo IBGE como um subgrupo de formação natural ou antropizado sujeito a fogo anual.

A segunda fonte de classificação “MODIS Agriculture” tem como objetivo mapear agricultura e agricultura + unidade de pastagem com base no comportamento de séries temporais ajustadas pela função do gradiente responsável por suavizar a série temporal atenuando o ruído. A terceira fonte de classificação “Pasture” foi originalmente implementada por Parente e Ferreira [9] que modificou a adição de dados da própria cobertura do solo como uma máscara para limitar a área de pixels do MODIS.

A aquisição dos dados de sensoriamento remoto para realizar o mapeamento de avaliação de uso e cobertura da área estudada, se deu através do processamento de imagens Landsat e MODIS disponíveis no catálogo do GEE para os anos de 2001, 2010 e 2017.

2.3 Matriz de Contingência

A matriz de contingência é uma tabulação cruzada simples dos rótulos de classe alocados pela classificação dos dados de sensoriamento remoto contra os dados referência para os sites de exemplo. A matriz de erro organiza os dados de amostra adquiridos de uma forma que resume os resultados e auxilia na quantificação de precisão e área. Na diagonal principal da matriz de erro estão destacadas as classificações corretas, enquanto os elementos fora da diagonal mostram erros de omissão e de comissão. As entradas de células e valores marginais da matriz de erro são fundamentais tanto para avaliação de precisão e estimativa da área [10].

3. RESULTADOS E DISCUSSÕES

Após a análise dos dados (Tabela 01), observou-se que no ano de 2001 a classe dominante era Savana Arborizada com 45% seguida da classe Savana (Sgf + Spf) com 34%, juntas formam 79% da área de estudo no período. As mesmas foram perdendo área nos anos de 2010 e 2017. Esse dado demonstra que essas classes apresentaram uma redução de mais de 30% em 2010/2017. Nas figura 4 é possível observar que as áreas de vegetação naturais foram reduzidas drasticamente.

Classes	2001	2010	2017
<i>Savanas (Sgf + Spf)</i>	34%	24%	20%
<i>Savana Arborizada</i>	45%	23%	23%
<i>Pastagem</i>	15%	34%	36%
<i>Outros</i>	6%	18%	16%
<i>Agricultura</i>	1%	2%	6%
Total Geral	100%	100%	100%

Tabela 1. Participação relativa das classes analisadas

As transições mostraram que em 2001 de 1, 773 Km² de Savana Florestada apenas 305 Km² foram convertidos em

Pasto e Agricultura, porém 952 Km² permaneceram Savana Florestada. Era o início do processo de consolidação da pecuária na região.



Figura 4. Pesquisa de campo, 2018

Analisando a classe Outros, verificou-se que ela possui grande intensidade dinâmica entre todas as classes, pudemos observar em campo que a classe Outros é composta principalmente por pasto sujo e formações vegetais do tipo Cerrado.

Uma característica importante que deve ser observada é o fato da presença cada vez maior da Soja na região sudeste do Pará (Figura 5). Muitas áreas de pastagem estão sendo substituídas pelas monoculturas de Soja e Milho e a produção de grãos em larga escala está se consolidando cada vez mais.



Figura 5. Pesquisa de Campo, 2018

Na tabela 2 estão apresentados os valores da matriz de contingência, gerada pela classificação em campo e pela imagem classificada, a partir dos quais foram obtidos os valores de Omissão, Inclusão.

Classificação	Agricultura	Outros	Pasto	Savana	Savana		Total	Omissão
					Arborizada			
Agricultura	3		2	1	3	9	0,33	
Outros	1	1			2	4	0,25	
Pasto		2	13	3	2	20	0,65	
Savana (Sgf + Spf)				9	2	11	0,82	
Savana Arborizada			1		16	17	0,94	
Total	4	3	16	13	25	61		
Inclusão	0,75	0,33	0,81	0,69	0,64			

Tabela 2. Matriz de contingência

A precisão global deste mapeamento foi de 68%. Talvez devido a influência do tamanho do pixel do MODIS, a classe

Savana Arborizada tenha apresentado um erro de omissão muito alto, da ordem de 0,94. A classe de pastagem, neste mapeamento foi confundida com a classe savana, possivelmente devido a sua similaridade espectral. A classe Agricultura apresentou um erro maior de inclusão 0,75, indicando que a área mapeada e maior do que a existente.

Os erros de omissão mostraram que do total de 25 pontos da classe Savana Arborizada apenas 16 foram classificadas como Savana Arborizada, os outros 9 pontos foram distribuídos entre as classes restantes não correspondendo à categoria na realidade. Dos 13 pontos da Savana 3 foram para a classe Pasto.

Os erros de inclusão correspondem aos pontos que foram devidamente incluídos em alguma classe. Na tabela 2 podemos observar que 3 pontos da classe Agricultura foram indevidamente incluídos na Classe Savana Arborizada, 2 na classe Pasto e 1 na Savana.

Do total de 4 pontos da classe Outros 2 foram inclusos na Savana Arborizada, e de 20 pontos do Pasto 3 foi para Savana, 2 para Savana Arborizada e 2 para a classe Outros. Dos 11 pontos da Savana, apenas 2 foram inclusos na Savana Arborizada e dos 17 pontos da Savana Arborizada 1 foi para a classe Outros.

4. CONCLUSÕES

A partir da classificação de uso e cobertura da terra em áreas de NF, foi possível observar a necessidade de monitoramento da região assim como as outras áreas da Amazônia são monitoradas atualmente.

A plataforma Google Earth Engine (GEE) através do algoritmo *Random Forest* mostrou-se uma ferramenta eficaz no mapeamento e classificação do uso da terra em áreas com formações não florestais. Com uma acurácia de 68%, o classificador apresentou mais pixels de confusão nas classes Outros, Savanas (Sgf + Spf) e Savana Arborizada por se tratar de áreas de diferentes fitofisionomias.

Outros estudos são necessários para avançar melhor nos resultados, porém por se tratar de uma área que causa muitas confusões de captura para o sensor, é um grande avanço analisar áreas de transição entre biomas por meio de dados de sensoriamento remoto.

A grande quantidade de dados em nuvem, a possibilidade de criar índices espectrais em segundos, e a variedade de algoritmos classificadores, permite ao usuário melhor compor seu set de classificação, portanto expõe o GEE como uma plataforma vasta em comandos, permissiva pela utilização e dinâmica para os usuários.

5. AGRADECIMENTOS

Os autores agradecem ao Programa de Pós-graduação em Ciências Ambientais (Universidade Federal do Pará, Campus Belém). Agradecem também à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). Ao Instituto Nacional de Pesquisas Espaciais (Centro Regional Amazônia

- CRA), pelo apoio e a FUNCATE (Fundação de Ciência, Aplicações e Tecnologias Espaciais) pela concessão da verba para realização do trabalho de campo.

6. REFERÊNCIAS

- [1] MALHI, Y. et al. Climate change, deforestation, and the fate of the Amazon. *Science*, v. 319, n. 5860, p. 169–172, 2008.
- [2] ROSSETTI, D. F.; TOLEDO, P. M. Biodiversity from a historical geology perspective: A case study from Marajo Island, lower Amazon. *Geobiology*, v. 4, n. 3, p. 215–223, 2006.
- [3] IBGE. Manual Técnico da Vegetação Brasileira. [s.l.: s.n.].
- [4] MAURYA, A. K.; TRIPATHI, S.; SONI, S. a Strategy of Forest / Non-Forest Cover Mapping of Achanakmar-Amarkantak Biosphere Reserve, Central India : Based on Remotely Sensed Imagery and Gis Data. v. 2, n. 4, 2013.
- [5] DINIZ, C. G. et al. DETER-B: The New Amazon Near Real-Time Deforestation Detection System. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 8, n. 7, p. 3619–3628, 2015.
- [6] CARVALHO, W. D., & MUSTIN, K. (2017). The highly threatened and little known Amazonian savannahs. *Nature Ecology and Evolution*. <https://doi.org/10.1038/s41559-017-0100>
- [7] FAPESPA, Fundação Amazônia de Amparo a Estudos e Pesquisas Estatísticas Municipais Paraenses: Santana do Araguaia. / Diretoria de Estatística e de Tecnologia e Gestão da Informação. – Belém, 2016. 54f.: il. Semestral, n. 1, jul. / Dez
- [8] WHITE, J. C. et al. Pixel-based image compositing for large-area dense time series applications and science. *Canadian Journal of Remote Sensing*, v. 40, n. 3, p. 192–212, 2014.
- [9] PARENTE, L.; FERREIRA, L. Assessing the spatial and occupation dynamics of the Brazilian pasturelands based on the automated classification of MODIS images from 2000 to 2016. *Remote Sensing*, v. 10, n. 4, 2018.
- [10] HEROLD, M. et al. Good Practices for Assessing Accuracy and Estimating Area of Land Change Good practices for estimating area and assessing accuracy of land change. [s.l.: s.n.]. v. 148.