

CLASSIFICAÇÃO DA VEGETAÇÃO DO PARQUE NACIONAL DA CHAPADA DAS MESAS, MARANHÃO, USANDO OBIA, MACHINE LEARNING E SOFTWARES LIVRES

Izadora Santos de Carvalho ¹, Tiago Massi Ferraz ², Thiago Sanna Freire Silva ³, Swanni T. Alvarado ⁴

¹Universidade Estadual do Maranhão - UEMA, Praça Gonçalves Dias, s/n – CEP 65800-000, Balsa-MA, e-mail: izadoraflorestal@gmail.com, ²Universidade Estadual do Maranhão-UEMA, Cidade Universitária Paulo VI, s/n – Tirirical – São Luís-MA e-mail: ferraztm@gmail.com, ³Universidade Estadual Paulista - UNESP, Avenida 24-A, nº 1515 – CEP 13506-900, Rio Claro-SP, ³e-mail: tsfsilva@rc.unesp.com, ⁴e-mail: swanni_ta@yahoo.es

RESUMO

Vários estudos têm demonstrado a importância do uso de algoritmos de aprendizado de máquina (*machine learning*, ML) para classificação de alvos a partir de imagens de sensoriamento remoto. Quando combinados a métodos de análise de imagens baseada em objetos (OBIA), resultados excelentes podem ser obtidos, especialmente em imagens com alta resolução espacial. O presente estudo objetivou classificar a vegetação do Parque Nacional da Chapada das Mesas (PNCM), utilizando métodos de OBIA e ML implementados em softwares de livre acesso. Para a classificação do PNCM foram utilizadas imagens da constelação Dove PlanetScope, da estação seca e chuvosa, e dados de elevação do modelo ALOS World 3D. Os procedimentos foram realizados usando a biblioteca RSGISLib la linguagem Python, a linguagem R e o software QGIS. Os seguintes procedimentos foram adotados para a realização do estudo: obtenção das imagens; segmentação; classificação; e filtragem e validação. O algoritmo utilizado para classificação foi o *Random Forests*. Para análise dos resultados foram gerados os índices *Kappa*, precisão global, e coeficientes de discordância quantitativa e de alocação, através da matriz de confusão, utilizando o pacote *rsacc* para o software R. Os resultados da classificação apontaram acurácia geral de 71,5%. O uso das ferramentas e softwares livres demonstraram-se promissores.

Palavras-chave — Código aberto, OBIA, cerrado, Random Forest, sensoriamento remoto.

ABSTRACT

Several studies have shown the importance of using machine learning algorithms for classifying targets in remote sensing images. When combined with Object-Based Image Analysis (OBIA) methods, they can yield excellent results, especially for high spatial resolution imagery. The present study aimed to classify the vegetation of the National Park of Chapada das Mesas (NPCM), using free, open access software. For the classification images from the Dove PlanetScope constellation were used, acquired at the dry and rainy seasons of the NPCM, together with elevation data from the ALOS World 3D model. The procedures were performed using the RSGISLib library of the Python language, the R language, and QGIS. The following procedures were

carried: segmentation; classification; and filtering and validation. The Random Forests classification algorithm was chosen, and for accuracy analysis, the Kappa, global precision, and quantity and allocation disagreement indexes were computed from the confusion matrix, using the rsacc R package. Classification results had 71.49% overall accuracy. The use of tools and free software has shown promise for future similar studies.

Key words — Open source, GEOBIA, savanna, Random forests, remote sensing.

1. INTRODUÇÃO

Os métodos de classificação de imagens são estudados e aprimorados constantemente. Com o avanço de softwares e a implementação de algoritmos modernos, trabalhar com imagens de satélite vem se tornando cada vez mais acessível, especialmente através dos softwares e linguagens open source [1]. Estes avanços permitem estudar a vegetação nas mais diversas escalas de tempo e espaço, (locais, regionais e globais), além de determinarem aspectos como os tipos de técnicas a serem aplicadas, softwares a serem utilizados, satélites e sensores, entre outros fatores [2].

Diversos estudos já demonstraram o sucesso dos métodos conhecidos como Análise de Imagens Baseada em Objetos (OBIA - *Object-Based Image Analysis*), especialmente para caracterização do uso e cobertura do solo e tipos de vegetações [3-6]. Com o crescente aumento de volume dados, e disponibilização de imagens multiespectrais e hiperespectrais, novas ferramentas para o processamento também são desenvolvidas [6]. Plataformas como *Google Earth Engine* (<https://earthengine.google.com>), *Anaconda* (<https://www.anaconda.com>), softwares como *QGIS* (<https://www.qgis.org>) e linguagens de código aberto (Python, R) possibilitam a execução de processos utilizando o conceito de *Object-Based Image Analysis* de forma livre e acessível [1, 6].

O presente estudo foi realizado utilizando pacotes de código aberto. Sendo um dos nossos objetivos difundir metodologias que usam ferramentas open source. A necessidade de um mapeamento mais detalhado das fisionomias do Parque Nacional da Chapada das Mesas (PNCM) impulsionou o presente estudo. Por não ter ainda um plano de manejo, necessita de pesquisas e ações que

possam subsidiar a implantação do mesmo, como o mapeamento detalhado das suas fisionomias.

Portanto, o objetivo deste trabalho foi realizar o mapeamento da vegetação do PNCM, utilizando o algoritmo de *machine learning* Random Forest, através do uso das bibliotecas RSGISLib e scikit-learn, acessadas por meio da linguagem Python. Buscou-se reponder as seguintes perguntas: a) qual a distribuição atual da cobertura vegetal no PNCM? b) qual a capacidade de bibliotecas e softwares de livre acesso em permitir a obtenção de mapas com boa acurácia através de métodos de OBIA e *machine learning*?

2. MATERIAIS E MÉTODOS

2.1. Local de estudo

A área selecionada para o estudo é a Unidade de Conservação de Proteção Integral Parque Nacional da Chapada das Mesas criado em 12 de dezembro de 2005, localizada no Sudoeste do estado do Maranhão, de coordenadas central $-07^{\circ} 10' 39.33''S$ e $-47^{\circ} 7' 29.45''O$. Localizado no Bioma Cerrado, o parque compreende os municípios de Estreito, Carolina e Riachão. A região é caracterizada por um clima Tropical Úmido, com duas estações bem definidas (seca e chuvosa).

O Relevo é predominantemente plano-ondulado, com a maior parte da área ocupada por relevo de chapada [7]. A área do PNCM possui uma drenagem rica, com muitos cursos d'água, sendo os principais o Rio Farinha e o Rio Itapecuru, além de possuir cerca de 400 nascentes [7].

2.2. Dados de entrada

Foram utilizadas imagens de alta resolução espacial da contelação Dove PlanetScope, operada pela empresa Planet (colocar link), adquiridas nas estações seca (2017-08-27) e chuvosa (2018-05-06). Estas imagens são multiespectrais e já corrigidas para efeitos atmosféricos, com 4 bandas (*blue*, *red*, *green* e *near-infrared*) e 3.5 metros de resolução espacial. Além das imagens PlanetScope, o *stack* utilizado para classificação da vegetação foi composto pelo modelo digital de elevação Alos World 3D (AW3D30) [8].

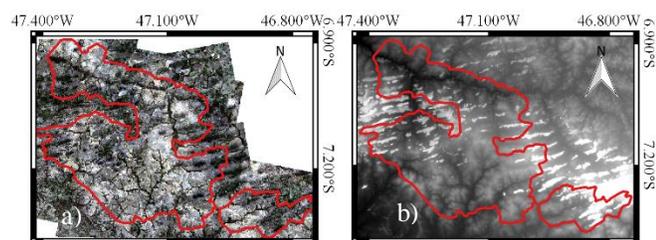


Figura 1. a) Mosaico das imagens PlanetScope; b) Imagem AW3D30 - MDE.

2.3. Segmentação

O algoritmo usado para a segmentação da imagem foi o de algoritmo de Shepherd, implementado na biblioteca RSGISLib [6]. Os parâmetros utilizados para executar o algoritmo foram: número de clusters $k = 10$, número mínimo de pixels por objeto = 600, e o limiar de distância (*distance threshold*) foi desabilitado, usando os valores padrão para os demais parâmetros. Para determinar os valores ideais dos parâmetros, diversos testes foram realizados até obter os agrupamentos mais homogêneos. A imagem utilizada para a segmentação foi composta por 8 bandas (4 da estação seca e 4 da estação chuvosa) para agregar o máximo de variações espectrais para a região classificada. O produto final dessa etapa foi uma imagem segmentada com estatísticas (mínimo, máximo, soma, média e desvio padrão) de todas as bandas das imagens (Figura 2) associadas aos objetos. Extraíram-se também, para os objetos as estatísticas referentes ao MDE.

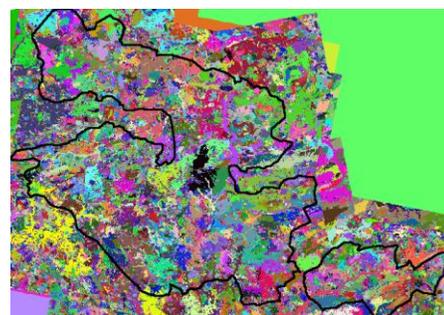


Figura 2. Imagem segmentada.

2.4. Classificação

No processo de classificação foram estabelecidas 9 classes de vegetação: cerrado, cerrado denso, cerrado aberto, cerrado rupestre, mata de galeria [9], agricultura, pastagem e água. Devido às características observadas em campo e nas imagens, a classe cerrado foi dividida em “cerrado planície” e “cerrado serrano”, diferenciados pela altitude. As amostras de treinamento das classes de fitofisionomias foram selecionadas por fotos e coordenadas de campo e por interpretação da imagem de alta resolução PlanetScope.

Para cada classe obteve-se um arquivo *shapefile* de treinamento, gerando um total de 791 amostras, distribuídas da seguinte forma: Água (n=179); Agricultura (n=50); Pastagem (n=45); Cerrado Planície (n=70); Cerrado Serrano (n=94); Cerrado Denso (n=65); Cerrado Aberto (n=75); Cerrado Rupestre (n=65); Mata de Galeria (n=148). A coleta das amostras foi realizada utilizando o software QGIS 2.18.19. Essas amostras foram divididas em aleatoriamente em amostras de treinamento e validação, na proporção de 70%-30%.

A classificação da vegetação foi realizada usando o algoritmo de *machine learning Random Forests*, da biblioteca Python scikit-learn, em conjunto com a biblioteca RSGISLib, ambos executados através de scripts em Python.

Todas as estatísticas das oito bandas espectrais me do MDE foram utilizadas como variáveis de entrada para a classificação. Para obter os melhores parâmetros para o classificador *Random Forest* foi executado uma otimização do tipo *GridSearch*, resultando no valor de ideal de 500 para o número de estimadores (árvores). O produto final dessa etapa foi uma imagem classificada com as 9 classes das fitofisionomias do PNCM.

2.5. Validação

A validação da classificação foi realizada na linguagem R [10] usando o pacote *rsacc - Accuracy Assessment for Remote Sensing Classification* (<https://github.com/EcoDyn/rsacc/>). Antes de executar a validação no R, a imagem produto da classificação passou por filtragem (*Majority Filter*) no QGIS, afim de eliminar *pixels* isolados e classificados incorretamente.

Para avaliação da qualidade da classificação foram utilizadas as análises dos coeficientes de discordância quantitativa e de alocação propostos por Pontius Jr e Millones [11], o índice *Kappa* e a precisão global, calculados a partir da matriz de confusão. A Figura 3 identifica todos os passos metodológicos utilizados.

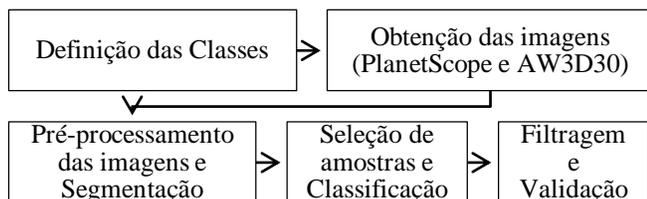


Figura 3. Fluxograma do processo de classificação das fitofisionomias do PNCM.

3. RESULTADOS E DISCUSSÃO

A Figura 4 mostra o produto final da classificação das fitofisionomias vegetais do PNCM, recortado a partir dos limites do parque. A imagem demonstra o mosaico de fitofisionomias existente no PNCM, destacando-se as classes de Cerrado Denso, Cerrado Aberto e Mata de Galeria.

Muitos estudos de classificação em regiões do Bioma Cerrado indicam dificuldades em mensurar classes como mata de galeria, cerrado denso, aberto e áreas antropizadas (agricultura e pastagem) [12, 13], muitas vezes devido à resolução espacial das imagens trabalhadas e métodos de classificação utilizados. Por exemplo, os autores Di Girolamo Neto et al. [13] encontraram problemas de confusão entre classes que exibem um gradiente de estrutura da vegetação entre elas, destacando que muitas vezes algumas classes devem ser unidas uma vez que a resposta espectral é parecida, dificultando assim a ação do classificador.

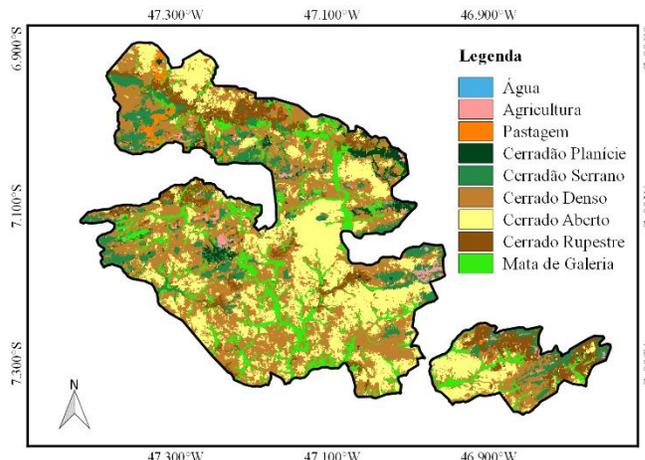


Figura 4. Mapa temático da classificação das fitofisionomias do PNCM.

No presente estudo as classes que apresentaram maior confusão foram Agricultura, seguida pelas classes Cerrado Planície e Pastagem. A partir da análise da matriz de confusão, observou-se a boa identificação da classe Cerrado Serrano, destacando assim o uso da banda de elevação [6] (Figura 5). Outros dados auxiliares para a classificação como: NDVIs, máscara de hidrografia e MNDWI, foram utilizados em estudos realizados por Fernandes et al., Ribeiro et al. e Nunes Roing [4,9,14], demonstrando outras possibilidades futuras para agregar informações adicionais para melhorar os resultados das classificações.

	Ág.	Agr.	Past.	C. Plan.	C. Serr.	C. Den.	C. Aber.	C. Rup.	Mt. de G.
Ág.	18	0	0	0	0	0	0	0	0
Agr.	0	7918	5087	7002	8027	0	0	0	0
Past.	0	1154	5091	0	0	0	0	0	0
C. Plan.	0	2181	981	4414	1378	0	0	0	218
C. Serr.	0	1361	174	9015	36844	685	0	0	5059
C. Denso	0	800	1387	0	187	44172	6873	3288	154
C. Aberto	0	0	0	0	0	4561	62251	2222	281
C. Rup.	1417	739	0	0	3	0	0	13027	479
Mata de G.	0	1020	268	8972	7577	0	1130	3271	44282

Figura 5. Matriz de confusão resultante do processo de validação da classificação. (Classes: Água = Ág.; Agricultura = Agr.; Pastagem = Past.; Cerradão Planície=C.Plan.; Cerradão Serrano = C.Serr.; Cerrado Denso = C.Dens.; Cerrado Aberto = C.Aber.; Cerrado Rupestre = C.Rup.; Mata de Galeria = Mt. de G.)

A classe água não obteve bons resultados, sendo pouco identificada, ocorrendo confusão entre a classe cerrado rupestre e água. Os cursos d’água do PNCM visíveis na imagem PlanetScope, como o rio Farinha, são caracterizados por possuírem diversas cachoeiras e pedrais de rochas. Esses fatores nos levam a acreditar que a semelhança da resposta espectral [13] dessas classes, seja o motivo de tal conflito.

Os valores de precisão global e índice *Kappa* foram de 71,49% e 65,99%, respectivamente. Considerados valores muito bons [14], levando em consideração o número de

classes e a resolução espacial das imagens utilizadas. Para um número menor de classes (5) e utilizando imagens de 30 metros de resolução, foi encontrado o valor de 84,4% de precisão global [13] para a mesma área de estudo. Contudo, os autores destacam a importância de inserir as classes florestais como a Mata de Galeria, que foi destacada na etapa de segmentação e diferem das formações savânicas [4, 9, 12]. A classificação da vegetação do PNCM teve o coeficiente de discordância de quantitativa de 16,59% e coeficiente de discordância de alocação de 11,92%.

Devido a heterogeneidade da área do PNMC, a classificação a partir de imagem apresenta dificuldades, o que é comum no Bioma Cerrado [12]. Contudo com o uso do método de classificação OBIA e das ferramentas e softwares acessíveis utilizados no presente estudo, os resultados para o desenvolvimento desse tipo de trabalho mostraram-se promissores.

5. CONCLUSÕES

O uso da imagem de alta resolução associada a dados de elevação aplicados para a classificação da vegetação do PNCM obteve bons resultados. O uso das ferramentas de código aberto, RSGISLib, Python, R e QGIS permitiu o desenvolvimento da metodologia OBIA com elevado desempenho, e sem custo, com valores muito bons de precisão global (71,49%) e índice *Kappa* (65,99%). Os resultados podem ser melhorados agregando mais dados no processo de classificação, diminuindo a ocorrência de conflitos entre classes que possuem respostas espectrais parecidas.

6. AGRADECIMENTOS

À Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, no âmbito do Edital CAPES/FAPEMA nº 29 /2017, pelo apoio financeiro, ao PPG em Agricultura e Ambiente – UEMA/Balsas e ao grupo de pesquisa Observatório de Dinâmicas Ecológicas – Ecodyn da UNESP de Rio Claro-SP. Agradecemos à Planet Labs por disponibilizarem as imagens PlanetScope através do Planet's Education and Research Program.

7. REFERÊNCIAS

- [1] Bunting, P.; Clewley, D.; Lucas, R.M. e Gillingham, S., The Remote Sensing and GIS Software Library (RSGISLib), *Computers and Geosciences*, v.62, pp.216-226, 2014.
- [2] Ponzoni, F.J. e Shimabukuro, Y.E., Sensoriamento remoto da vegetação, Oficina de Textos, São Paulo, pp.165, 2012.
- [3] Dronova, I.; Gong, P.; Clinton, N.E.; Wang, L.; Fu, W.; Qi, S. e Liu, Y., Landscape analysis of wetland plant functional types: The effects of image segmentation scale, vegetation classes and classification methods, *Remote Sensing of Environment*, v.127, pp.357-369, 2012.
- [4] Fernandes, R.; Nunes, G.M. e Silva, T.S.F., Classificação orientada a objetos aplicada na caracterização da cobertura da terra no Araguaia, *Pesquisa Agropecuária Brasileira*, v.47, n.9, pp.1251-1260, 2012.
- [5] Evans, T.L. e Costa, M., Landcover classification of the Lower Nhecolândia subregion of the Brazilian Pantanal Wetlands using ALOS/PALSAR, RADARSAT-2 and ENVISAT/ASAR imagery, *Remote Sensing of Environment*, v.128, n.9, pp.118-137, 2013.
- [6] Clewley, D.; Bunting, P.; Shepherd, J.; Gillingham, S.; Flood, N.; Dymond, J.; Lucas, R.; Armston, J.; Moghaddam, M., A python-based open source system for geographic object-based image analysis (GEOBIA) utilizing raster attribute tables, *Remote Sensing*, v.6, n.7, pp.6111-6135, 2014.
- [7] Lima, L.P.; Tchaicka, L.; Avelar, J.J.C. e Cunha, A.N.C., Plano operativo de prevenção e combate aos incêndios florestais do Parque Nacional da Chapada das Mesas, Ministério do Meio Ambiente-MMA, pp.18, 2007.
- [8] Tadono, T.; Nagai, H.; Ishida, H.; Oda, F.; Naito, S.; Minakawa, K. e Iwamoto, H. Initial Validation of the 30 m-mesh Global Digital Surface Model Generated by ALOS PRISM. The International Archives of the Photogrammetry, *Remote Sensing and Spatial Information Sciences, ISPRS*, v.XLI-B4, pp.157-162, 2016.
- [9] Ribeiro, J.F. e Walter, B.M.T., As principais fitofisionomias do Bioma Cerrado. In Cerrado: ecologia e flora (S.M. Sano, S.P. Almeida & J.F. Ribeiro, eds.). Embrapa Cerrados, Planaltina. pp.151-212, 2008.
- [10] R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL Available at: <http://www.R-project.org>, 2018.
- [11] Pontius Jr, R.G.; Millones, M. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, v.32, pp.4407-4429, 2011.
- [12] Sano, E.E.; Rosa, R.; Brito, J.L.S.; Ferreira, L.G., Land cover mapping of the tropical savanna region in Brazil, *Environment Monitoring and Assessment*, v.166, pp.113-124, 2010.
- [13] Di Girolamo Neto, C.; Fonseca, L.M.G.; Valeriano, D.M.; Neves, A.K.; Körting, T.S., Desafios na classificação automática de fitofisionomias do Cerrado brasileiro com base em mapas de referência na escala 1:250.000. In: Simpósio Brasileiro de Sensoriamento Remoto, 18, 2017, Santos, SP. *Anais...*, pp. 6647-6654, 2017.
- [14] Nunes, J.F. e Hoig, H.L., Análise e mapeamento do uso e ocupação do solo da bacia do Alto do Descoberto, DF/GO, por meio de classificação automática baseada em regras e lógica nebulosa, *Revista Árvore*, v.39, n.1, pp.25-36, 2015.