

AVALIAÇÃO DE ÍNDICES ESPECTRAIS E DO ALGORITMO RANDOM FOREST PARA A DETECÇÃO DE MUDANÇAS DA COBERTURA DO SOLO NO CERRADO BRASILEIRO

Álvaro Salgado Araújo Silva Pinto¹, Eduarda Martiniano de Oliveira Silveira², Thainá Guimarães Rocha³, Luíza Imbroisi Ferraz Cunha⁴, Fausto Weimar Acerbi Júnior⁵

Departamento de Ciências Florestais (DCF), Universidade Federal de Lavras, Brasil; ¹alvarosalgado96@yahoo.com.br; ²dudalavras@hotmail.com; ³thaina.g.rocha@gmail.com; ⁴luiza.1302@gmail.com; ⁵fausto@dcf.ufla.br

RESUMO

O objetivo deste trabalho foi analisar a capacidade de 7 índices de vegetação em detectar e diferenciar mudanças na cobertura do solo em uma cena Landsat inserida no Cerrado Brasileiro, para isto foi utilizado o algoritmo Random Forest. Foram geradas as imagens diferença entre os anos de 2006 a 2017, estas imagens foram segmentadas e posteriormente foram coletadas 50 amostras por classe e ano das classes queimada, desmatamento, regeneração e não mudança. A partir destas amostras extraiu-se valores mínimos, médios, máximos e desvios padrão para treinamento do algoritmo Random Forest. Como resultado, verificou-se que todos os índices contribuíram para a acurácia global de 92,37%, comprovando sua alta capacidade em diferenciar vegetação de não vegetação. Os índices NBR e NBR2 foram os mais significativos ao diferenciar queimadas de desmatamentos, sendo que a alta separabilidade destes índices também foi comprovada em outros estudos no Cerrado.

Palavras-chave — Sensoriamento remoto, índices de vegetação, desmatamentos, queimadas, áreas em regeneração.

ABSTRACT

The objective of this work was to analyze 7 vegetation indices in order to detect and differentiate land cover changes in the Brazilian savannas, also known as the Cerrado. We calculated the difference images from Landsat scenes of the years 2006 to 2017. The difference images were segmented and 50 samples per class (burned, deforestation, regeneration and no changed areas) per year were collected. Then, we extracted from each sample the minimum, average, maximum and standard deviation values for training the Random Forest algorithm. The results showed the importance of all indices to achieve the overall accuracy of 92.37%, proving their high capacity to differentiate vegetated from non-vegetated areas. The NBR and NBR2 were the most significant indices for differentiation of burned areas and deforestation. Similar

results were also reported by other authors studying land cover changes in the Cerrado.

Key words — Remote sensing, vegetation indices, deforestation, burned, regeneration areas.

1. INTRODUÇÃO

O cerrado Brasileiro consiste no segundo maior bioma da América do Sul e está localizado em sua maioria na região Centro-Oeste, mas também compreende partes das regiões Norte, Nordeste e Suldeste do país. Ocupa 21% do território nacional [1], e abriga mais de 7000 espécies arbóreas, sendo considerada a mais diversificada savana do mundo [2]. A exploração do bioma iniciou-se na década de 40 com a criação do Projeto de Colonização nos Cerrados, porém grande parte da conversão da vegetação nativa deveu-se as políticas de expansão agropecuária na década de 70 [3]. Portanto, amparadas no lema “Integrar para não entregar” as políticas de expansão agropecuária foram efetivas, tornando o Cerrado um polo da agricultura e da pecuária. Contudo, o saldo negativo é uma área desmatada de 272.198,8 km² entre os anos de 2001 a 2017 [4].

Além disso, as queimadas causadas por uso antrópico do fogo ou queda de raios são frequentes no bioma e podem produzir incêndios impossíveis de serem controlados pelo homem [5], devido ao grande acúmulo de biomassa seca e de palha durante a estação seca. O sensoriamento remoto constitui-se como a única alternativa viável para estudar o funcionamento de grandes biomas como o Cerrado [6]. Entretanto, as mudanças constantes e a sazonalidade se tornam desafios para o mapeamento e monitoramento da cobertura do solo [7], uma vez que as mudanças fenológicas podem ser detectadas como desmatamento.

O objetivo desse estudo foi avaliar a acurácia da classificação de imagens utilizando o algoritmo Random Forest [8] e índices espectrais derivados do satélite Landsat, para classificar áreas de desmatamento, queimadas, regeneração e áreas sazonais (não mudança) no Cerrado Brasileiro.

2. MATERIAIS E MÉTODOS

2.1. Caracterização da área de estudo

A área de estudo (Figura 1) está localizada na cena 219/71 do satélite Landsat, na região norte do estado de Minas Gerais (MG), Brasil. A área abrange 19 municípios inseridos na bacia hidrográfica do Rio São Francisco, o bioma predominante é o Cerrado.

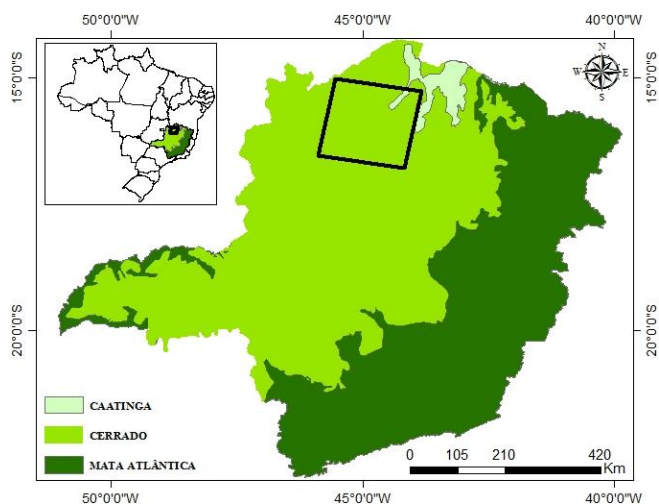


Figura 1. Área de estudo (Cena 219/71), contendo as delimitações dos biomas Brasileiros no Estado de Minas Gerais e sua localização no país.

2.2. Aquisição de imagens

As imagens foram adquiridas a partir do portal USGS *Earth Resources Observation and Science* (EROS), provenientes dos satélites LANDSAT 5 e LANDSAT 8 (resolução espacial de 30 metros), órbita 219 e ponto 71, entre os anos de 2006 a 2017 (imagens anuais), exceto do ano 2012, onde não há imagens disponíveis para esse satélite. Juntamente com as bandas foram obtidos 7 índices espectrais (Tabela 1).

Índice	Fórmula	Referência
EVI	$G \frac{\text{NIR}-\text{Red}}{\text{NIR}+C1*\text{Red}-C2*\text{Blue}+L}$	(9)
SAVI	$\frac{\text{NIR}-\text{Red}}{\text{NIR}+\text{Red}} (1+1)$	(10)
MSAVI	$\frac{2\text{NIR}+1-\sqrt{(2\text{NIR}+1)^2-8(\text{NIR}-\text{Red})}}{2}$	(11)
NBR	$\frac{\text{NIR} - \text{SWIR2}}{\text{NIR} + \text{SWIR2}}$	(12)
NBR2	$\frac{\text{SWIR1} - \text{SWIR2}}{\text{SWIR1} + \text{SWIR2}}$	(12)
NDMI	$\frac{\text{NIR} - \text{SWIR1}}{\text{NIR} + \text{SWIR1}}$	(13)
NDVI	$\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$	(14)

Tabela 1. Índices de vegetação utilizados no estudo.

2.3. Processamento

O primeiro processo foi a geração das imagens diferenças de cada banda e índice espectral. As imagens diferenças contendo as bandas espectrais foram utilizadas para a segmentação de imagens utilizando o algoritmo de multiresolução [15] no *software* eCognition. Os segmentos obtidos foram exportados para o *software* ArcGis, onde foram coletadas amostras de treinamento (50 objetos por classe e por ano), contemplando as classes: desmatamento, queimadas, regeneração e não mudança (áreas com sazonalidade) (Figura 2).

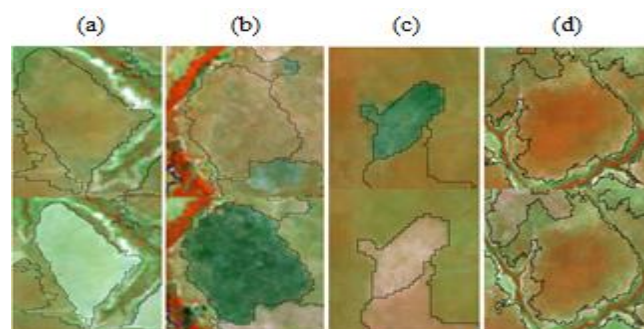


Figura 2. Exemplo das classes: (a) desmatamento; (b) queimada; (c) regeneração e (d) não mudança.

De cada segmento em cada ano, foram extraídos os valores mínimos, médios, máximos e desvios padrão de cada índice espectral para treinamento do algoritmo Random Forest [16] utilizando o *software* R [17].

2.4. Random Forest

O Random Forest é um algoritmo de aprendizagem de máquina, que pode ser empregado em classificações ou regressões. Para isso cria árvores de decisão e as combina, visando predições com maior acurácia e estabilidade [18]. Cerca de 70% das amostras foram utilizadas para treinamento enquanto os 30% restantes foram utilizados para validação. Os valores das acurácias (acurácia global, acurácia do produtor e acurácia do usuário) foram obtidos através da matriz de confusão [19] e índice kappa [20].

Dois parâmetros foram definidos para a execução do Random Forest: o número de árvores de decisão a serem geradas (Ntree) e o número de variáveis a serem selecionadas e testadas por árvore (Mtry). Quinhentas árvores foram construídas e o Mtry foi igual a 1 [21]. Após a classificação, o IncMse % (incremento no erro médio quadrático) foi obtido para verificação do índice espectral que possibilita a melhor separabilidade entre as classes.

3. RESULTADOS

A classificação obtida com o algoritmo Random Forest apresentou resultados satisfatórios na separação entre as

classes desmatamento, queimadas, regeneração e não mudança, atingindo acurácia global de 92,37%. As maiores acuracidades do usuário foram encontradas para as classes regeneração e desmatamento, sendo 96,75% e 95,46%, respectivamente. Dentre as classes analisadas, o desmatamento foi a que apresentou menor acuracidade de produtor (90,83%) (Tabela 2).

Classes	Ac. Usuário (%)	Ac. Produtor (%)
Desmatamento	95,46	90,83
Queimada	85,14	93,71
Não mudança	88,28	93,39
Regeneração	96,75	94,90
Ac. Global (%)	92,37	

Tabela 2. Matriz de confusão.

Os sete índices de vegetação utilizados no estudo contribuíram para a alta acurácia da classificação. Entre eles o mais significativo foi o NBR2, especificamente nas estatísticas desvio padrão e média, com %IncMse de 62,26% e 44,22%, respectivamente (Figure 3).

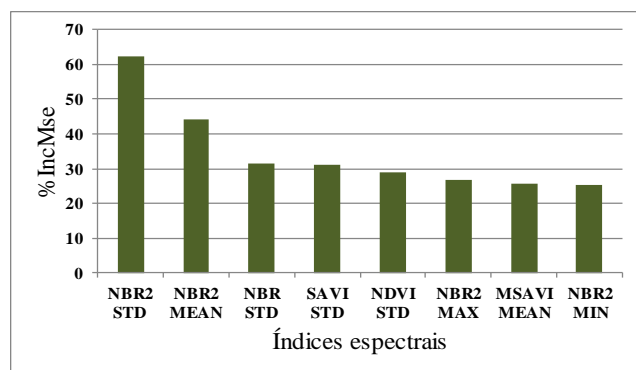


Figura 3. % IncMse de cada índice espectral utilizado na classificação.

Também foi calculado o índice kappa que é tido como uma técnica multivariada discreta que define se uma matriz de classificação é significativamente diferente de outra [22]. Para esse estudo o índice kappa foi de 88,63%, que de acordo com a literatura [23] trata-se de uma classificação quase perfeita.

4. DISCUSSÃO

A grande sazonalidade e heterogeneidade de fitofisionomias em biomas como o Cerrado faz com que haja diversos padrões espectrais para as classes de interesse, o que influi diretamente nos resultados de classificações provenientes de algoritmos como o Random Forest.

Os índices de vegetação vêm sendo utilizados com intuito de melhorar a separabilidade entre classes, por realçarem as diferenças espectrais entre pares de bandas, a partir da análise das curvas de assinatura espectral de

determinados alvos [24]. O princípio básico desses índices consiste em diferenciar vegetação de não vegetação, reunindo as informações espectrais [25]. Isto faz com que os índices de vegetação sejam importantes para classificar e diferenciar queimadas e desmatamentos no Cerrado. Dos 7 índices utilizados, o NBR2 obteve maior êxito na discriminação entre polígonos de queimada e desmatamento. Isto também pôde ser notado em outros trabalhos que tiveram regiões do Cerrado Brasileiro como área de estudo. Por exemplo, num estudo com o objetivo de separar queimadas de outros alvos utilizando a mesma cena deste trabalho, MIRBI e NBR2 foram os índices que resultaram em maior separabilidade [26]. O índice NBR apresentou resultados significativos para o mesmo propósito na cena 220/68, onde os erros de omissão e comissão foram inferiores a 5% [27]. Em outro estudo utilizando as cenas 219/70 e 221/67, os índices NBR2 e MIRBI foram os que apresentaram maior diferenciação entre áreas queimadas e não queimadas [28].

Especificamente na detecção de áreas queimadas os índices de vegetação que apresentaram maior capacidade de discriminação foram o NBR2, NDVI e W [29]. Considerando o mapeamento de áreas em regeneração foi o NBR o índice que apresentou melhor distinção entre as fisionomias vegetais [30].

5. CONCLUSÕES

Por meio da validação dos valores estatísticos no algoritmo Random Forest, foi possível observar que todos os 7 índices utilizados no estudo contribuíram para a alta separação entre as amostras de queimada e desmatamento, sendo que os índices NBR2 e NBR foram os mais significativos neste sentido.

Conclui-se que os índices de vegetação possuem grande potencial no que se diz respeito ao monitoramento de queimadas e desmatamentos no Cerrado brasileiro. A partir deles é possível detectar, separar e quantificar com maior precisão estas mudanças, o que poderá facilitar e automatizar o aumento e a manutenção das bases de dados existentes para este bioma.

6. REFERÊNCIAS

- [1] Klink, C. A.; Machado, R. B., "A conservação do cerrado brasileiro", *Megadiversidade*, v. 1, n. 1, pp. 147-155, 2005.
- [2] Sawyer, D., "População, meio ambiente e desenvolvimento sustentável no cerrado.", In: Hogan, D.J.; Cunha, J.M.P.; Carmo, R.L. (org.). *Migração e ambiente no Centro-Oeste*, UNICAMP: PRONEX, Campinas, pp. 279-299, 2002.
- [3] Miziara, F. & Ferreira, N. C., "Expansão da fronteira agrícola e evolução da ocupação e uso do espaço no Estado de Goiás: Subsídios à política ambiental." In: Ferreira, L.G. (org.). *A encruzilhada socioambiental - Biodiversidade, economia e sustentabilidade no cerrado*, CEGRAF-UFG, Goiânia, pp. 67-75, 2008.

- [4] INPE: Incrementos de desmatamento anuais no Cerrado Brasileiro, disponível em: <<http://www.terrabrasilis.dpi.inpe.br/dashboard/deforestation/biomes/cerrado/increments/>>. Acesso em: outubro/2018
- [5] Coutinho, L.M. O bioma do Cerrado. In: KLEIN, A. L. (Org.), “Eugen Warming e o Cerrado brasileiro um século depois.”, UNESP, São Paulo, p. 77-91, 2002
- [6] Asner, G. P.; Knapp, D. E.; Cooper, A. N.; Bustamante, M. C.; Orlander, L. P. “Ecosystem structure throughout the Brazilian Amazon from Landsat observations and automated spectral unmixing.”, *Earth Interactions*, v. 9, n. 7, pp. 31, 2005.
- [7] Sano, E. E.; Rosa, R.; Brito, J. L. S.; Ferreira, L. G., “Mapeamento semidetalhado do uso da terra do Bioma Cerrado.” *Pesquisa Agropecuária Brasileira*, v. 43, n. 1, pp. 153-156, 2008.
- [8] Breiman L, “Random forests”. *Machine Learning*, v. 45, n. 1, pp. 5–32, 2001.
- [9] Huete A, Didan K, Miura H, et al., “Overview of the radiometric and biophysical performance of the MODIS vegetation indices.”, *Remote Sensing of Environment*, v. 83, n. 1, pp. 195–213, 2002.
- [10] Huete AR, “A soil-adjusted vegetation index (SAVI).”, *Remote Sensing of Environment*, v. 25, n. 3, pp. 295–309, 1988.
- [11] Rondeaux G, Steven M, Baret F, “Optimization of soil-adjusted vegetation indices”. *Remote Sensing of Environment*, v. 55, n. 2, pp. 95–107, 1996.
- [12] Miller JD, Thode AE, “Quantifying burn severity in a heterogeneous landscape with a relative version of the delta Normalized Burn Ratio (dNBR).”, *Remote Sensing of Environment*, v. 109, n. 1, pp. 66–80, 2007.
- [13] Wilson EH, Sader SA, “Detection of forest harvest type using multiple dates of Landsat TM imagery.”, *Remote Sensing of Environment*, v. 80, n. 3, pp. 385–396, 2002.
- [14] Tucker CJ, “Red and photographic infrared linear combinations for monitoring vegetation.”, *Remote Sensing of Environment*, v. 8, n. 2, pp. 127–150, 1979.
- [15] Baatz, M. and Schape, A., “Multiresolution Segmentation: An Optimization Approach for High Quality Multi-Scale Image Segmentation.”, In: Strobl, J., Blaschke, T. and Griesbner, G., Eds., *Angewandte Geographische Informations-Verarbeitung XII*, Wichmann VerlagI, Karlsruhe, pp. 12-23, 2000.
- [16] Liaw A, Wiener M, “Classification and Regression by random forest.” *R News*, v. 2, pp. 18-22, 2002.
- [17] R Core Team, “A language and environment for statistical computing.”, *R foundation for Statistical Computing*, Vienna, 2014.
- [18] DeVries, B., A. K. Pratihast, J. Verbesselt, L. Kooistra, M. Herold, “Characterizing Forest Change Using Community-Based Monitoring Data and Landsat Time Series.”, *PLoS ONE*, 11 (3), pp. 1–25, 2016.
- [19] Congalton, R.G., “A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data.”, *Remote Sensing of Environment*, v. 37, n. 1, pp. 35-46, 1991.
- [20] Cohen, J., “A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*.”, v. 20, n. 1, pp. 37-46, 1960.
- [21] Silveira, E.M.O., I.T. Bueno, F.W. Acerbi-Júnior, J.M. Mello, J.R.S. Solforo, M.A. Wulder, “Using Spatial Features to Reduce the Impact of Seasonality for Detecting Tropical Forest Changes from Landsat Time Series.” *Remote Sensing*, v. 10, n. 6, pp. 808, 2018.
- [22] Bishop, Y.M.M., et al., “Discrete Multivariate Analysis: Theory and Practice.”, *Mass MIT Press, Cambridge*, pp. 18-37, 1975.
- [23] LANDIS, J.R.; KOCH, G.G. The measurement of observer agreement for categorical data. *Biometrics*, v. 33, n. 1, pp. 159-174, 1977
- [24] Marcussi, A. B.; Bueno, C. R. P.; Miqueloni, D. P.; Arraes, C. L., “Utilização de índices de vegetação para os sistemas de informação geográfica.”, *Caminhos da geografia Jaboticabal*, v. 11, n. 35, pp. 41-53, 2010.
- [25] Liu, W.T.H., “Aplicações de sensoriamento Remoto.”, *UNIDERP, Campo Grande*, pp. 908, 2007.
- [26] Pereira, A.A.; De Carvalho, L.M.T.; Libonati, R.; Acerbi Júnior, F.W.; Morelli, F., “Avaliação de nove índices espectrais quanto a separabilidade entre queimadas e diferentes alvos”, *Simpósio Brasileiro de Sensoriamento Remoto*, v. 17, pp. 3105-3012, 2015.
- [27] Barroso, A.G., E.E. Sano, D.M. Freitas, “Identificação de desmatamentos recentes no cerrado utilizando as técnicas de diferença de imagens e índices de queimada por diferença normalizada.”, *Rev. Bras. Cartografia*, v. 69, n. 7, pp. 1397-1409, 2017.
- [28] Pereira, A.A.; Fabrício Rodrigues Teixeira, R.L.; Melchiori, E.A.; Carvalho, L.M.T., “Avaliação de índices espectrais para identificação de áreas queimadas no cerrado utilizando dados LandSat TM.”, *Rev. Bras. Cartogr.*, v. 68, n. 8, pp. 1665–1680, 2016.
- [29] Melchiori, A. E.; Cândido, P. A.; Libonati, R.; Morelli, F.; Setzer, A.; Jesus, S. C.; Fonseca, L. M. G.; KORTING, T. S “Spectral indices and multitemporal change image detection algorithms for burned area extraction in the Brazilian Cerrado.” *Simpósio Brasileiro de Sensoriamento Remoto*, v. 17, pp. 643-650, 2015.
- [30] Resende, F. C.; Soares, T. B. O.; Santos, P. R.; Pereira, G., “Análise de Índices Espectrais para Estimativa de Áreas de Regeneração Florestal no Parque Nacional Chapada Das Mesas.”, *Revista Territorium Terram*, v. 3, n. 5, pp. 95-104, 2015.