

MAPEAMENTO DE FRAGMENTOS FLORESTAIS POR MEIO DE TÉCNICA DE MINERAÇÃO DE DADOS

Rosane Gomes da Silva ^{1,3}, Ronie Silva Juvanhol ², Alexandre Rosa dos Santos ³

¹Instituto Federal de Educação, Ciência e Tecnologia do Norte de Minas Gerais/ IFNMG, Campus Araçuaí, Fazenda Meio Pé da Serra – BR 367, km 47, s/n 39600-000, Araçuaí, MG, Brasil. rosanegomes.eaa@gmail.com; ²Universidade Federal do Piauí/ UFPI, Av. Manoel Gracindo, s/n, 64900-000, Bom Jesus, PI, Brasil, roniejuvanhol@gmail.com; ³Universidade Federal do Espírito Santo/ UFES, Programa de Pós graduação em ciências florestais, Av. Governador Lindemberg, 316, 29550-000, Jerônimo Monteiro, ES, Brasil. alexandre.santos@pq.cnpq.br

RESUMO

O objetivo deste trabalho foi delimitar os fragmentos de mata atlântica no PARNA Caparaó e parte da sua zona de amortecimento, utilizando metodologia de mineração de dados, associado ao uso de imagens do sensor MODIS. Foram utilizadas imagens de NDVI, EVI, IAF, albedo, PPB e FPAR. Em cada imagem foram selecionadas 3000 amostras aleatórias, para as classes floresta e não floresta. Foi testado o algoritmo de árvore de decisão J48, que apresentou um valor de índice Kappa de 0,74. Após a poda da árvore de decisão, as regras foram implementadas para classificar as áreas de floresta e não floresta. Assim, obteve-se o mapa de fragmentos florestais para a área de estudo, em que o índice Kappa calculado foi de 0,70. Conclui-se que a utilização da árvore de decisão permitiu a extração de informações úteis referentes a áreas de floresta em imagens de sensoriamento remoto.

Palavras-chave — MODIS, Árvore de decisão, J48, PARNA Caparaó, Índice KAPPA.

ABSTRACT

The objective of this study was to map the Atlantic forest fragments at PARNA Caparaó and part of its buffer zone, using data mining methodology, associated to the use of MODIS sensor images. Images of NDVI, EVI, IAF, albedo, PPB and FPAR were used. In each image, 3000 random samples were selected for the forest and non-forest classes. The decision tree algorithm J48 was tested, which presented a Kappa index value of 0.74. After the pruning of the decision tree, the rules were implemented to classify the forest and non-forest areas. Thus, the map of forest fragments was obtained for the study area, in which the calculated Kappa index was 0.70. It is concluded that the use of the decision tree allowed the extraction of useful information regarding forest areas in remote sensing images.

Key words — MODIS, Decision tree, J48, PARNA Caparaó, Kappa index.

1. INTRODUÇÃO

Os ecossistemas florestais são áreas de grande relevância ambiental, pois a vegetação é responsável por regular o balanço de energia, ciclo da água e ciclo biogeoquímico na superfície terrestre por meio da fotossíntese, respiração, transpiração, albedo da superfície e rugosidade [1]. O bioma mata atlântica encontra-se entre os ecossistemas mais ameaçados [2], apresentando elevado grau de fragmentação florestal [2, 3].

Dessa forma, o monitoramento da cobertura vegetal torna-se importante como auxílio à conservação florestal, destacando-se dentre os métodos modernos o sensoriamento remoto, pela capacidade de fornecer informação de forma sinóptica, abrangendo áreas extensas e com alta frequência de aquisição [4, 5]. Dentre os produtos de sensoriamento remoto, imagens do sensor *Moderate Resolution Imaging Spectroradiometer* (MODIS) têm sido amplamente utilizadas nas mais diversas aplicações ambientais. A principal vantagem é a disponibilidade de 44 produtos, demandando menor necessidade de processamentos para utilização das imagens, associada à elevada resolução espectral, que registra informações em 36 faixas do espectro eletromagnético [6].

É possível combinar produtos de diferentes características da vegetação de floresta, uma vez que a resposta neste tipo de ambiente apresenta padrões desconhecidos, mas que podem ser detectados pela associação entre diferentes informações [7, 8]. A árvore de decisão é um método de mineração de dados que se destaca no contexto de classificação de imagens, pois proporciona um aumento da acurácia e permite classificar conjuntos distintos de dados [9]. É um método em que a classificação é obtida de forma explícita, simplificando a interpretação dos resultados [10].

Neste contexto, o objetivo deste trabalho foi delimitar os fragmentos de mata atlântica no Parque Nacional do Caparaó e parte da sua zona de amortecimento, utilizando classificador de árvore de decisão J48 e imagens de moderada resolução espacial do sensor MODIS.

2. MATERIAIS E MÉTODOS

2.1. Área de estudo

A área de estudo abrange o Parque Nacional do Caparaó (PARNA Caparaó) e a parte da sua zona de amortecimento no estado do Espírito Santo, definida como uma faixa de 10

km de distância a partir dos limites do Parque, de acordo com o que foi estabelecido em seu plano de manejo. Localiza-se na divisa do sudoeste do estado do Espírito Santo com Minas Gerais, entre os paralelos 20°12' e 20°42' de latitude Sul e os meridianos 41°52' e 41°38' de longitude Oeste (Figura 1).

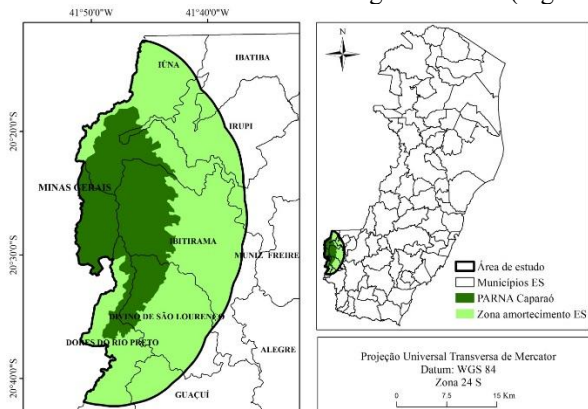


Figura 1. Localização da área de estudo.

O PARNA Caparaó possui uma área aproximada de 317,03 km², sendo pertencente ao domínio de Mata Atlântica [11]. Na área do entorno, a vegetação nativa é caracterizada pela alta fragmentação florestal, o que é comum de ocorrer neste bioma [2, 3].

2.2. Dados de sensoriamento remoto

Foram utilizados quatro produtos do sensor MODIS, referentes ao tile h14v11, correspondendo a seis imagens que representam características da vegetação, sendo estas: índices de vegetação: *Normalized Difference Vegetation Index* (NDVI) e *Enhanced Vegetation Index* (EVI), Índice de área foliar (IAF), albedo, produtividade primária bruta (PPB) e fração de radiação fotossinteticamente ativa (FPar). Foi padronizado o ano de 2007 para obtenção das mesmas e escolhidos os meses de fevereiro ou abril, considerando datas em que a floresta apresenta maior vigor vegetativo, pois coincide com o período chuvoso na região. As informações das imagens e produtos utilizados encontram-se melhor detalhadas na Tabela 1.

Tabela 1. Produtos MODIS utilizados e suas principais características

Produto	Imagens	Resolução espacial	Resol. temporal	Data
MCD15A2H	FPar e IAF	500 m	8 dias	15/04/2007
MOD13Q1	EVI e NDVI	250 m	16 dias	22/04/2007
MOD17A2H	PPB	500 m	8 dias	10/02/2007
MCD43A3	Albedo	500 m	16 dias	18/02/2007

O NDVI é um índice de vegetação que traduz o contraste gerado pela absorção de pigmentos vermelhos pela clorofila e a alta reflectância do mesófilo na faixa do infravermelho próximo [12]. O EVI foi desenvolvido para otimizar o sinal

da resposta da vegetação, melhorando a sensibilidade em áreas com maiores densidades de biomassa [6].

A Produtividade Primária Bruta está relacionada às trocas de energia e massa em superfícies vegetais e ao balanço de carbono, sendo considerada como a soma da fotossíntese realizada por todas as folhas em escala ecossistêmica [13]. Já a FPAR é definida como a fração da radiação fotossinteticamente ativa incidente (400-700nm) absorvida pelos elementos verdes de um dossel de vegetação.

O Índice de área foliar (IAF) é definido como a área projetada pelas folhas por unidade de área de solo [14]. O Albedo ou refletividade da superfície de ondas curtas é um parâmetro importante no orçamento de energia de superfície [15], que varia de acordo com a entrada de calor solar, separação e realocações entre o sistema terra-atmosfera [16].

As imagens MODIS são disponibilizadas na projeção sinusoidal e referenciadas para o Norte dos Estados Unidos, sendo necessárias as correções da projeção, sistema de referência e localidade. Para isso, foi utilizado o software *Modis Reprojection Tools* (MRT), de acesso livre e download disponível no site da *National Agency and Space Administration* (NASA). Inicializado o aplicativo, os dados preenchidos foram: imagens de entrada, quadrante da área de estudo, local de saída dos dados, formato dos dados de saída como GeoTiff, sistema de coordenadas geográficas e sistema de projeção WGS 84.

Após essa etapa, as imagens foram recortadas, tendo como base arquivo vetorial em formato *shape*, representativo da área do Parque Nacional do Caparaó e a parte capixaba de sua zona de amortecimento.

2.3. Classificação das imagens

Foram selecionadas 3000 amostras aleatórias para o conjunto de treinamento para as classes: “Floresta” e “Não Floresta”, e extraídos os valores pontuais dessas amostras para cada uma das seis imagens consideradas no estudo. Esses dados foram exportados para uma tabela no *software excel* e convertidos para um formato de arquivo *.csv*, para serem implementados ao *software* de mineração de dados Weka 3.6.

Foi utilizado o algoritmo J48, com as configurações padrão do *software*. Para a avaliação do algoritmo J48, foi considerado o número de acertos, representado pelo valor de índice Kappa, e a complexidade da árvore gerada, pois é importante que o método seja aplicável para estudos posteriores. O índice kappa define uma medida de acurácia de acordo com a proporção em que uma área é corretamente classificada com base em uma referência, sendo calculado conforme a equação 1.

$$K = \frac{\theta_1 - \theta_2}{1 - \theta_1} \quad (1)$$

$$\text{Onde, } \theta_1 = \frac{\sum_{i=1}^c x_{ii}}{n} \text{ e } \theta_2 = \frac{\sum_{i=1}^c x_i}{n^2}$$

O índice Kappa varia de 0 (uma classificação muito fraca) até 1 (uma classificação excelente) (Tabela 2).

Tabela 2. Avaliação da classificação de acordo com o índice Kappa

Índice Kappa	Qualidade da classificação
0	Muito fraca
0-0,2	Fraca
0,2-0,4	Razoável
0,4-0,6	Boa
0,6-0,8	Muito boa
0,8-1	Excelente

A árvore de decisão foi podada, retirando as regras não aplicáveis aos dados. As regras obtidas dessa maneira foram utilizadas para a classificação da área nas classes “Floresta” e “Não Floresta”.

De posse do arquivo vetorial representativo da delimitação das áreas de florestas, fotointerpretado em aerofoto na escala 1:1500, para o ano de 2007 [17], foi possível obter a matriz de confiança para a classificação gerada para toda a área de estudo. Primeiramente, o arquivo vetorial foi rasterizado e atribuídos os valores 1 para “Floresta” e 2 para “Não Floresta”. Então, foi gerada a matriz de confiança e, assim, avaliado o resultado final da classificação.

A avaliação levou em consideração medidas derivadas da matriz de confiança, tais como o índice Kappa (K), mencionado anteriormente e descrito de acordo com a equação 1; o Erro de Omissão (EO), definido como os pixels representativos da categoria “Floresta” que foram classificados como “Não Floresta” (Equação 2); o Erro de Comissão (EC), definidos como os pixels correspondentes à classe “Não Floresta” e classificados erroneamente como “Floresta” (Equação 3), e o Viés (V); definido como a proporção entre os pixels classificados como “Floresta” e os pixels que realmente pertencem à essa classe (Equação 4).

$$OE = \frac{X_{21}}{X_{11} + X_{21}} \quad (2)$$

$$CE = \frac{X_{12}}{X_{11} + X_{12}} \quad (3)$$

$$V = \frac{X_{11} + X_{12}}{X_{11} + X_{21}} \quad (4)$$

Onde X representa os elementos da matriz de confiança gerada.

3. RESULTADOS E DISCUSSÃO

A árvore J48 apresentou um valor de índice Kappa satisfatório, de 0,74 e necessitou de um número pequeno de poda. As regras que não fizeram sentido foram: $EVI \leq 3212$ e $EVI \leq 5186$: Floresta, pois maiores valores de índice de vegetação correspondem à maior massa vegetal, característica de floresta. Realizada a poda da árvore, a estrutura de regras implementada foi a seguinte (Figura 2).

Após a implementação do algoritmo de árvore de decisão para classificação, obteve-se o mapa de fragmentos florestais para o ano de 2007, referente à área de estudo. Na Figura 3 pode-se observar o mapa real (a), obtido por meio de fotointerpretação em tela e considerado como a realidade, em

contraste com o mapa classificado pela árvore de decisão (b). Os erros de Omissão, correspondentes aos pixels de floresta classificados como Não Floresta, e de comissão, correspondentes aos pixels de Não Floresta classificados como Floresta, podem ser visualizados na Figura 3c.

```

Albedo <= 128: Floresta
Albedo > 128
| NDVI <= 7921: NF
| NDVI > 7921
| | PPB <= 87: Floresta
| | PPB > 87
| | | NDVI <= 8309: NF
| | | NDVI > 8309: Floresta
    
```

Figura 2. Estrutura da árvore de decisão após a poda

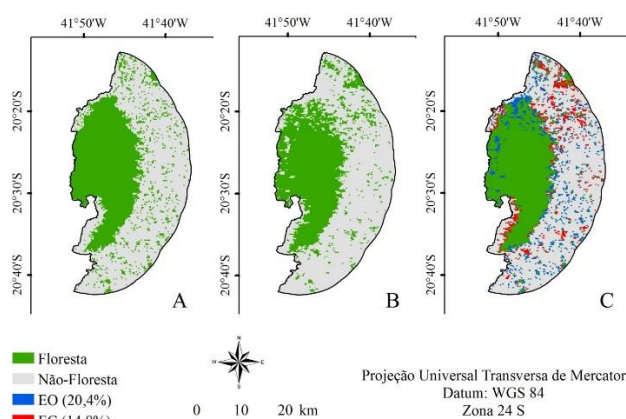


Figura 3. Mapas obtidos: (a) pela fotointerpretação (real), (b) por meio da implementação do algoritmo de árvore de decisão j48 e (c) espacialização dos erros de omissão e comissão

Com base na matriz de confiança (Tabela 4), foram calculados os índices para avaliação da classificação K, EO, EC e Viés. O índice Kappa encontrado foi de 0,70, classificado como um valor bom. [18] utilizaram técnicas de mineração de dados associada ao uso de imagens de sensoriamento remoto e conseguiram um valor de índice Kappa de 0,76 na classificação. As imagens, nesse caso, apresentaram uma resolução espacial de 30 m e foram associadas à composição IHS e cálculo de textura na área de estudo.

Tabela 4. Matriz de confiança obtida para a classificação das imagens

Classes	Floresta	Não Floresta
Floresta	5285	919
Não Floresta	1357	8386

Os erros de Omissão e Comissão encontrados foram de 0,2 e 0,15; respectivamente, indicando que a maior parte do erro esteve associada à omissão de pixels representativos de

áreas de floresta na classificação e uma menor parte associada a pixels classificados como “Não Floresta”, que na realidade pertenciam à classe “Floresta”. O valor encontrado para o viés foi de 0,93, associado a uma pequena superestimação da classe “Floresta” pela classificação. É importante considerar que os baixos valores encontrados para os erros de omissão e comissão, e o valor do viés próximo a 1, indicam uma boa confiabilidade da classificação, pois são valores baixos de erros e uma superestimação muito pequena.

Considerando a utilização de imagens de sensoriamento remoto com resolução espacial de 250 e 500 m, o resultado encontrado foi bastante satisfatório. Além disso, a região do estudo encontra-se em alto grau de fragmentação florestal, em que os fragmentos são muitos e de tamanho pequeno. Assim, a superestimação dos valores pode estar associada à detecção de áreas florestais de forma correta pelos dados de sensoriamento remoto, que puderam ser computados considerando apenas tamanhos de pixel de no mínimo 250 m.

[8] compararam diferentes métodos para classificação de uso da terra por meio da utilização de dados do sensor MODIS associado à técnicas de mineração de dados. Os autores obtiveram um índice Kappa de 0,84 ao combinarem métricas de fenologia e bandas espectrais do sensor. No caso do presente estudo, ficou comprovado que a combinação dos diferentes produtos do sensor MODIS, permite a distinção da classe de floresta das demais e possui grande potencial para ser utilizado em trabalhos para classificação de uso da terra com diferentes coberturas florestais.

5. CONCLUSÕES

A associação dos diferentes produtos do sensor MODIS possibilitou a distinção das áreas de floresta das demais, em especial as imagens de NDVI, albedo e produtividade primária bruta. Simultaneamente, a utilização da árvore de decisão permitiu a extração de informações úteis referentes a áreas de floresta em imagens de sensoriamento remoto.

6. REFERÊNCIAS

- [1] Jackson, R.B.; Randerson, J.T.; Canadell, J.G.; Anderson, R.G.; Avissar, R.; Baldocchi, D.D.; Bonan, S.P. Protecting climate with forests. *Environment Research Letters*, v. 3, pp. 44006, 2008.
- [2] Ribeiro, M.C., Metzger, J.P., Martensen, A.C., Ponzoni, F.J., Hirota, M.M. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation*. V. 142, pp. 1141–1153. 2009.
- [3] Juvanol, R.S., Fiedler, N.C., Santos, A.R., Pirovani, D.B., Louzada, F.L.R. de O., Dias, H.M., Tebaldi, A.L.C. Análise Espacial de Fragmentos Florestais: Caso dos Parques Estaduais de Forno Grande e Pedra Azul, Estado do Espírito Santo. *Floresta e Ambiente*. v.18, pp. 353–364. 2011.
- [4] Richards, J. A. Remote sensing digital image analysis: An introduction. Springer-Verlag, Heiderberg. 2 ed. 340 p. 1993.

- [5] Xie, Y., Sha, Z., Yu, M. Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology*. v. 1, pp. 9–23. 2008.

- [6] Justice, C.O.; Townshend, J.R. G.; Vermote, E. F.; Masuoka, E.; Wolfe, R. E.; Saleous, N.; Roy, D. P.; Morisette, J. T. An overview of MODIS Land data processing and product status. *Remote Sensing of Environment*, v. 83, n. 1-2, pp. 3–15, 2002.

- [7] Carvalho Júnior, O.A., Coelho, M.A.N., Martins, É. de S., Gomes, R.A.T., Couto Junior, A.F., Oliveira, S.N. de, Santana, O.A. Mapeamento da vegetação na floresta atlântica usando o classificador de árvore de decisão para integrar dados de sensoriamento remoto e modelo digital de terreno. *Revista Brasileira de Geofísica*. v. 26, pp. 331–345. 2008.

- [8] Zhou, F., Zhang, A., Townley-Smith, L. A data mining approach for evaluation of optimal time-series of MODIS data for land cover mapping at a regional level. *ISPRS Journal of Photogrammetry and Remote Sensing*. v. 84, pp. 114–129. 2013.

- [9] Hansen MC, Roy DP, Lindquist E, Adusei B, Justice CO, Altstatt A. A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. *Remote Sensing of Environment*. v. 112, n.5, pp. 2495–2513. 2008.

- [10] Stiglic G, Kocbek S, Pernek I, Kokol P. Comprehensive decision tree models in bioinformatics. *Plos One*. v. 7, n. 3, pp. 1–13. 2012.

- [11] IBAMA – Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis. Roteiro metodológico para a elaboração de planos de ação para implementação e gerenciamento de unidades de conservação de uso indireto. IBAMADIREC: Brasília. 24 p. 1993

- [12] Bannari A, Morin D, Bonn F, Huete AR. A review of vegetation indices. *Remote Sensing Reviews*. v. 13, n (1–2), pp. 95–120. 1995.

- [13] Chapin III FS, MATSON PA, MOONEY HA. Principles of Terrestrial Ecosystem EcologySpringer-Verlag: . Berlin. 2002.

- [14] Ross J. The radiation regime and architecture of plant stands. 3rd ed. Springer Netherlands: Dordrecht. 1981.

- [15] Cao Y, Liang S, Chen X, He T. Assessment of Sea Ice Albedo Radiative Forcing and Feedback over the Northern Hemisphere from 1982 to 2009 Using Satellite and Reanalysis Data. *Journal of Climate*. v. 28, n. 3, pp. 1248–1259. 2015.

- [16] Trenberth KE, Fasullo JT, Kiehl J. Earth’s Global Energy Budget. *Bulletin of the American Meteorological Society*. v. 90, n. 3, pp. 311–323. 2009.

- [17] Santos AR, Chimalli T, Peluzio JBE, Silva AG, Santos GMADA, Lorenzon AS et al. Influence of relief on permanent preservation areas. *Science of The Total Environment*. v. 541, pp. 1296–1302. 2016.

- [18] Fernandes P.J.F., Girão R.S., Silva L.C. Técnicas de processamento digital de imagens e mineração de dados aplicadas ao mapeamento de fragmentos florestais a partir de imagens ETM+. *Caderno de Estudos Geoambientais*. v. 4, n. 1, pp. 23–36. 2013.