

GOOGLE EARTH ENGINE COMO INSTRUMENTO DE CLASSIFICAÇÃO USO E COBERTURA DO SOLO NO MUNICÍPIO DE CASCAVEL, PARANÁ, BRASIL

Willyan Ronaldo Becker¹, Lucas Eduardo de Carvalho², Jerry Adriani Johann³, Laíza Cavalcante de Albuquerque Silva¹, Alex Paludo¹, Jonathan Richetti¹, Eduardo Matias Rambo¹

¹Universidade Estadual do Oeste do Paraná, Programa de Pós-Graduação em Engenharia Agrícola, Cascavel, Paraná, Brasil, {willyanbecker@hotmail.com; laiza.cavalcante@hotmail.com; paludo.alex@hotmail.com; j_richetti@hotmail.com; eduardomrambo@hotmail.com}; ²Universidade Estadual do Oeste do Paraná, Graduando em Engenharia Agrícola, Cascavel, Paraná, Brasil, {lucasak_47carvalho@hotmail.com}; ³Universidade Estadual do Oeste do Paraná, Docente no Programa de Pós-Graduação em Engenharia Agrícola, Cascavel, Paraná, Brasil, {jerry.johann@unioeste.br}

RESUMO

Este trabalho teve como objetivo a realização do mapeamento do uso e cobertura do solo para o município de Cascavel, localizado na mesorregião Oeste do estado do Paraná. Foi utilizado imagens do satélite Landsat-8 e processadas na plataforma web de processamento de dados geoespaciais Google Earth Engine, a qual se trata de um ambiente de processamento em nuvem, com objetivo de fazer análise de dados ambientais em grande escala e com recursos computacionais amplos, utilizando os servidores Google para processamento. Foram calculados os índices de vegetação SR, NDVI, EVI, SAVI e NDWI, além do desvio padrão do perfil de NDVI. Os resultados mostram que a classe predominante no município de Cascavel é Agricultura, com 59,02% da área territorial do município, mostrando a importância da agricultura para o Cascavel e região. A segunda maior classe foi Mata Nativa, com 25,94%, seguida de Pastagem (7,61%), Reflorestamento (3,20%) e Corpo Hídrico (0,10%).

Palavras-chave — Mineração de Dados, Random Forest, Exatidão Global.

ABSTRACT

The objective of this work was the mapping of land use and land cover for the municipality of Cascavel, located in the western mesoregion of the state of Paraná. Images from the Landsat-8 satellite were used and processed on the geospatial data processing web platform Google Earth Engine, which is a cloud processing environment, with the purpose of analyzing large-scale environmental data with large computational resources, using Google servers for processing. Vegetation indices SR, NDVI, EVI, SAVI and NDWI were calculated, in addition to the standard deviation of the NDVI profile. The results show that the predominant class in the municipality of Cascavel is Agriculture, with 59.02% of the territorial area of the municipality, showing the importance of agriculture for Cascavel and region. The second largest class was Forest, with 25.94%, followed by Pasture (7.61%), Reforestation (3.20%) and Water (0.10%).

Key words — Data Mining, Random Forest, Global Accuracy.

1. INTRODUÇÃO

O uso e cobertura do solo é uma das informações mais relevantes na interpretação de uma imagem de satélite, pois permite a identificação e distribuição espacial do tipo de cobertura que reveste o solo [1] conforme sua natureza. Tal conhecimento é de grande importância ao planejamento por contribuir na orientação da ocupação da paisagem, respeitando sua capacidade de suporte e/ou sua estabilidade/vulnerabilidade [1,2].

O sensoriamento remoto desponta como uma ferramenta para monitorar a dinâmica de ocupação pela rápida execução e aquisição de dados, e pelo menor custo atrelado ao processo se comparado aos métodos com observação *in loco*. Para tanto, a resolução das imagens dos satélites é um dos fatores fundamentais para a acurácia do mapeamento, uma vez que a interpretação, tanto do profissional quanto do algoritmo classificador, está em função da distinção entre os alvos terrestres [3].

Neste sentido, o processo de Descoberta de Conhecimento em Base de Dados (*Knowledge Discovery From Data* – KDD), especificamente algoritmos de mineração de dados, possibilita encontrar relações ou padrões ocultos em bases de dados, com a finalidade de converter dados em conhecimento [4,5], facilitando assim o processo de mapeamento.

Todavia, monitorar grandes áreas por um longo intervalo de tempo eleva consideravelmente o recurso computacional necessário para o processamento de dados geoespaciais. A plataforma em nuvem Google Earth Engine (GEE), possibilita a análise de dados ambientais em grande escala por utilizar os servidores da Google. Além disso, o GEE proporciona ao usuário uma abordagem altamente interativa para o desenvolvimento de algoritmos e um ciclo rápido de testes e melhorias para o processamento de dados em larga escala [6].

Assim, o objetivo deste trabalho foi conhecer a proporção de uso do solo entre as classes Mata, Corpo Hídrico, Reflorestamento, Pastagem e Agricultura para o município de Cascavel, Paraná, Brasil, com imagens do satélite Landsat-8 para o ano de 2017 utilizando como ferramenta o GEE.

2. MATERIAIS E MÉTODOS

As etapas do processo executado são apresentadas na Figura 1, onde em (a) é apresentada as etapas realizadas na plataforma Google Earth Engine.

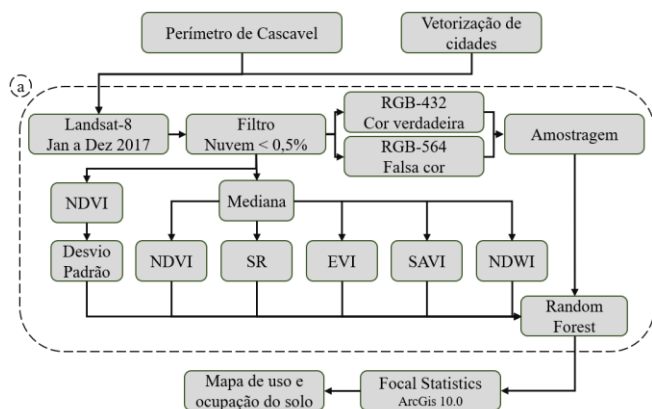


Figura 1. Fluxograma geral de trabalho para obtenção do uso e cobertura do solo para o município de Cascavel. Em (a) as etapas realizadas no Google Earth Engine.

A área de estudo compreendeu o município de Cascavel, localizado na mesorregião Oeste do estado do Paraná. O município conta com uma área territorial de 2.100,831 km² [7], e está inserido em uma região com elevada importância na produção agrícola.

O clima de Cascavel, segundo a classificação de Köppen, se enquadra no tipo Cfa [8]. O município está presente no Terceiro planalto paranaense e apresenta áreas com ocorrência de Latossolos e Nitossolos, que são solos profundos e bem drenados, como também Neossolos, que são solos rasos em áreas de maior declividade [9].

A vetorização do perímetro urbano, distritos, vilas e demais áreas construídas foi executada software Google Earth Pro. A vetorização teve como intuito reduzir a classificação e contabilização dessas áreas nas demais classes de interesse.

O arquivo vetorial de perímetro do município foi importado na plataforma GEE e utilizado para a delimitação e recorte das imagens Landsat-8. As imagens possuem resolução espacial de 15 (pancromática, banda 8), 30 metros (bandas 1 a 7) e 100 metros (bandas 10 e 11) e temporal de 16 dias. Como a computação do Google Earth Engine se dá na menor resolução das imagens de entrada [6], os resultados podem ser obtidos na escala de 15 metros, porém foi definido a escala de 30 metros.

Para que a presença de nuvens não influenciasse nas métricas a serem calculadas a partir das bandas multispectrais, foi definido o percentual máximo de nuvem de 0,5% nas imagens. Ao todo, foram utilizadas 11 imagens no ano de 2017. As bandas adotadas foram B1 (0.43 - 0.45 µm), B2 (0.45 - 0.51 µm), B3 (0.53 - 0.59 µm), B4 (0.64 - 0.67 µm), B5 (0.85 - 0.88 µm), B6 (1.57 - 1.65 µm), B7 (2.11 - 2.29 µm), B8 (0.52 - 0.90 µm), B9 (1.36 - 1.38 µm), B10 (10.60 - 11.19 µm) e B11 (11.50 - 12.51 µm).

O script desenvolvido em javascript calcula a mediana de cada banda e com isto determinou-se os índices de vegetação: *Simple Ratio* (SR – Equação 1), *Normalized Difference Vegetation Index* (NDVI – Equação 2), *Enhanced Vegetation Index* (EVI – Equação 3), *Normalized Difference Water Index* (NDWI – Equação 4) e *Soil-Adjusted Vegetation Index* (SAVI, Equação 5). A partir do perfil espectro-temporal do índice NDVI foi calculado o desvio padrão (SD – Equação 6) para as 11 imagens do período.

$$SR = \frac{NIR}{RED} \quad (1)$$

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (2)$$

$$EVI = 2,5 \frac{(NIR + C_1 * RED - C_2 + BLUE + L)}{(NIR + RED)} \quad (3)$$

$$NDWI = \frac{NIR - GREEN}{NIR + GREEN} \quad (4)$$

$$SAVI = \frac{NIR - RED}{NIR + RED + L} * (1 + L) \quad (5)$$

$$SD = \sqrt{\frac{\sum_{i=1}^N (X - \bar{X})^2}{N}} \quad (6)$$

em que:

NIR: refletância na faixa do infravermelho próximo;

RED: refletância na faixa do vermelho;

GREEN: refletância na faixa do verde;

C1: coeficiente de correção de aerossol (6);

C2: coeficiente de correção de aerossol (7,5);

L: fator de correção para a interferência do solo (1 para o EVI e 0,5 para o SAVI).;

X: valor do NDVI do pixel da imagem;

$\bar{X}$: média de NDVI dos pixels para a mesma localização;

N: número total de imagens.

A seleção dos dados de treinamento foi realizada pela RGB-432 (cor verdadeira) e RGB-564 (falsa-cor). Ao todo, foram coletadas 345 amostras, subdivididas em 100 amostras de Mata Nativa, 35 de Corpos Hídricos, 75 de Reflorestamento, 35 de Pastagem e 100 de Agricultura.

O mapeamento do uso e cobertura do solo foi efetuado com o algoritmo de classificação Random Forest (RF). O RF cria várias árvores de decisão, treinadas por meio da seleção aleatória de 75% dos dados de treinamento e 25% para a validação cruzada da árvore gerada [5]. Foi definido o total de 100 árvores para a classificar a área de estudo.

Na etapa de avaliação da qualidade da classificação foram sorteados 50 pontos aleatórios em cada classe estudada, totalizando 250 pontos e, junto com imagens de alta resolução do Google Earth, foram confirmados ou não a real classificação destes. Assim, foi gerada a matriz de confusão e, a partir desta, foram calculadas métricas de acurácia: Acurácia de Produtor (AP), Acurácia de Usuário (AU), Exatidão Global (EG), índice Kappa (K) e índice Tau (T).

3. RESULTADOS E DISCUSSÃO

A Figura 2 traz a distribuição espacial do uso e cobertura do solo para o município de Cascavel, Paraná, Brasil. Por análise visual, percebe-se que a região norte de Cascavel é predominantemente agrícola. A região leste do município é dividida entre agricultura, mata nativa e reflorestamento. Já a parte sul de Cascavel é dividida entre as classes de agricultura, mata nativa, reflorestamento e pastagem. Também se observa que a região sul apresenta talhões com dimensões menores, devido a sua característica de relevo acidentado, enquanto a porção norte é menos irregular, a partir de observações de dados obtidos da missão *Shuttle Radar Topography Mission* (SRTM) [10].

A área urbana obtida por meio de vetorização manual apresentou um percentual de 4,14% da área total do município. A área ocupada por Agricultura correspondente a 59,02% da área territorial de Cascavel, sendo a classe com maior representatividade no mapeamento, mostrando a importância do agronegócio para o município. Segundo [11], a área de soja em Cascavel para o ano-safra 2016/2017 foi de 92.000 hectares e 4.550 hectares de milho, sendo estas as duas principais culturas de verão, totalizando 96.550

hectares. No mapeamento por sensoriamento remoto foi encontrado um total de 135.973 hectares de agricultura, valor 40% superior aos dados oficiais.

A área de Mata nativa foi a segunda maior classe estimada, com 25,94% da área de Cascavel. A cobertura do município ocupada por Corpos hídricos é de 0,10%, principalmente composto por açudes. A maior cobertura de corpo hídrico da região é o lago municipal de Cascavel, porém este foi retirado junto com a área urbana, por estar no interior desta. A área de Pastagem foi estimada em 7,61% e a área de Reflorestamento em 3,20%.

Analisando a AU os valores foram de 0,90 para Mata nativa, 1,0 para Água, 0,98 para Reflorestamento, 0,92 para Pastagem e 0,96 para Agricultura. As principais omissões foram nas classes Reflorestamento em Mata e Agricultura em Pastagem, igualmente justificáveis devido à similaridade espectro-temporal entre os alvos. Tais resultados são superiores quando comparados com os obtidos por [12] na região de Jacarta, Indonésia. No trabalho, os autores encontraram AU de 0,817, 0,783 e 0,547 para as classes de agricultura, área urbana e mata, respectivamente. Já para os AP, foram de 0,704, 0,846 e 0,716.

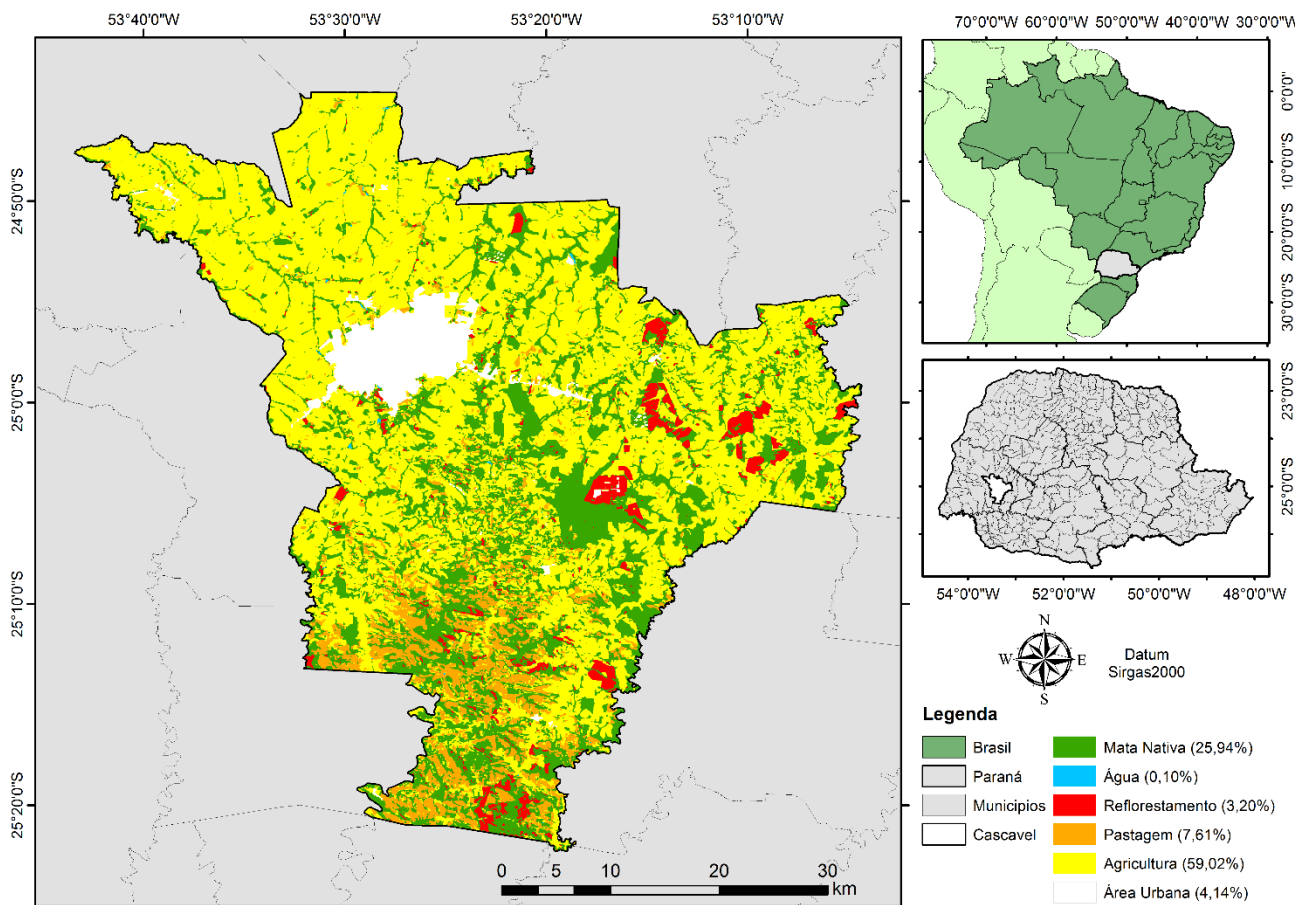


Figura 2. Mapa de uso e cobertura do solo para o município de Cascavel, ano de 2017.

As AP foram de 0,94 para Mata, 0,96 para Água, 0,94 para Reflorestamento, 0,98 para Pastagem e 0,94 para Agricultura. As principais inclusões foram de Pastagem em Mata, Mata em Reflorestamento e Pastagem em Agricultura, justificáveis devido a similaridades espectro-temporal entre os alvos.

A Exatidão Global calculada foi de 95,2%, com índice Kappa de 0,94 e índice Tau de 0,94. Tais valores são considerados excelentes [13,14].

5. CONCLUSÕES

O mapeamento de uso e cobertura do solo com a utilização da mediana das bandas e dos índices de vegetação SR, NDVI, EVI, SAVI e desvio padrão do perfil espectro-temporal de NDVI do satélite Landsat-8 se mostrou como uma alternativa com excelentes resultados, podendo-se agrupar os alvos no solo conforme as classes de interesse.

A utilização da plataforma Google Earth Engine no mapeamento permitiu a realização de várias simulações e visualização rápida dos resultados por conta do processamento em nuvem no qual a plataforma se baseia. A partir deste script, é possível agora a realização de mapeamentos de uso e ocupação de solo em outros municípios, bastando apenas sejam realizadas a seleção de amostras dos alvos de interesse que contemplem toda a região de estudo e que estas estejam bem distribuídas.

6. AGRADECIMENTOS

Os autores agradecem a Universidade Estadual do Oeste do Paraná (UNIOESTE), a Pós-Graduação em Engenharia Agrícola (PGEAGRI) e ao Núcleo de Pesquisa GeoScience pela infraestrutura ofertada e apoio técnico-científico e a Fundação Araucária (FA) e o Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

7. REFERÊNCIAS

- [1] SEMA, Secretária de Estado do Meio Ambiente e Recursos Hídricos. “Bacias Hidrográficas do Paraná”. 2a edição. Curitiba/PR: 2013.
- [2] Leite, E. F.; Rosa, R. “Análise do Uso, Ocupação e Cobertura da Terra na Bacia Hidrográfica do Rio Formiga, Tocantins”. OBSERVATORIUM: Revista Eletrônica de Geografia, volume (4), número (12), páginas (90–106), 2013.
- [3] Formaggio, A. R.; Sanches, I. D. “Sensoriamento Remoto em Agricultura”. São José dos Campos - SP: Eficina de Textos, 2017.

- [4] Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. “Knowledge Discovery and Data Mining: Towards a Unifying Framework”. In: Knowledge Discovery and Data Mining Portland OR. p. 82–88.

- [5] Tan, P. N.; Steinbach, M.; Kumar, V. “Introdução ao data mining: Mineração de dados”. Rio de Janeiro/RJ: Ciência Moderna, 2009.

- [6] Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. “Google Earth Engine: Planetary-scale geospatial analysis for everyone”. Remote Sensing of Environment, volume (202), páginas (18–27), 2017.

- [7] IBGE, Instituto Brasileiro de Geografia e Estatística. “Brasil em Síntese”. Disponível em: <<https://cidades.ibge.gov.br/brasil/panorama>>. Acesso em: 9 ago. 2018.

- [8] Aparecido, L. E. de O.; Rolim, G. de S.; Richetti, J.; Souza, P. S. de; Johann, J. A. “Köppen, Thornthwaite and Camargo climate classifications for climatic zoning in the State of Paraná, Brazil”. Ciência e Agrotecnologia, volume (40), número (4), páginas (405–417), 2016.

- [9] Bhering, S. B.; Santos, H. G. dos; Manzatto, C. V.; Bognola, I.; Fasolo, P. J.; Carvalho, A. P. de; Potter, R. O.; Curcio, G. R.; Aglio, M. L. D.; Souza, J. S. de; Chaffin, C. E.; Junior, W. de C. “Mapa de Solos do Estado do Paraná”. Embrapa, Empresa Brasileira de Pesquisa Agropecuária, 2007.

- [10] Farr, T.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; Seal, D.; Shaffer, S.; Shimada, J.; Umland, J.; Werner, M.; Oskin, M.; Burbank, D.; Alsdorf, D. “The shuttle radar topography mission”. Reviews of Geophysics, volume (45), páginas (1–33), 2007.

- [11] IBGE, Instituto Brasileiro de Geografia e Estatística. “Banco de dados agregados”. Disponível em: <<http://www.sidra.ibge.gov.br>>. Acesso em: 8 ago. 2018.

- [12] Tsutsumida, N.; Comber, A. J. “Measures of spatio-temporal accuracy for time series land cover data. International Journal of Applied Earth Observation and Geoinformation”, volume (41), páginas (46–55), 2015.

- [13] Foody, G. M. “Status of land cover classification accuracy assessment”. Remote Sensing of Environment, volume (80), número (1), páginas (185–201), 2002.

- [14] Congalton, R. G.; Green, K. “Assessing the Accuracy of Remotely Sensed Data: Principles and Practices”. 2nd edition. ed. Boca Raton, FL, USA: CRC Press, 2009.