

PROCESSAMENTO E CLASSIFICAÇÃO DE IMAGENS LANDSAT-8 PARA USO E OCUPAÇÃO DO MUNICÍPIO DE TUPÃSSI - PR UTILIZANDO O GOOGLE EARTH ENGINE

Eduardo Matias Rambo¹, Laiza Cavalcante de Albuquerque Silva¹, Willyan Ronaldo Becker¹, Jonathan Richetti¹, Joseane Bortolini¹, Jerry Adriani Johann², Erivelto Mercante²

¹Universidade Estadual do Oeste do Paraná, Programa de Pós-Graduação em Engenharia Agrícola, Cascavel, Paraná, Brasil, {eduardorambo@hormail.com; laiza.cavalcante@hotmail.com; willyanbecker@hotmail.com; j_richetti@hotmail.com; josi_bortolini@hotmail.com}; ²Universidade Estadual do Oeste do Paraná, Docente no Programa de Pós-Graduação em Engenharia Agrícola, Cascavel, Paraná, Brasil, {jerry.johann@hotmail.com; eriveltomercante@yahoo.com.br}

RESUMO

O uso de imagens de satélites associada a técnicas de interpretação e classificação possibilitam o mapeamento de áreas, como também a análise do grau e forma do uso e ocupação do solo. Este trabalho teve como objetivo classificar o uso e ocupação do município de Tupãssi, Paraná, utilizando cenas do satélite Landsat-8/OLI pelos classificadores *Cascade K-Means* e *Naive Bayes* na plataforma Google Earth Engine e pela vetorização manual. Os alvos contidos no município foram subdivididos em cinco classes de interesse: soja, milho, água, vegetação (floresta e reflorestamento) e edificações. As análises de acurácia dos mapeamentos foram elaboradas a partir da vetorização do município. Obteve-se que a cultura predominante no município foi a soja, ocupando 85% da área, seguida do alvo vegetação com 13% de área. O classificador *K-means* obteve um EG 87,86% e Kappa 0,56 e o classificador *Naive Bayes* um EG de 88,59% e Kappa 0,58.

Palavras-Chave – *Naive Bayes, Cascade K-Means, Google Earth Engine.*

ABSTRACT

Classification techniques in satellite images enable mapping land cover usage and quantification of area with rapidly time process using Google Earth Engine platform. Hence, this work aims to classify Tupãssi municipality, in Paraná State, Brazil, using Landsat-8/OLI images by the classifiers Cascade K-Means, Naive Bayes and manual vectorization in GEE. Five targets were used: soybean, corn, water, forest and urban areas as input. The dominant target was soybean (85% of area), followed by 13%. The results shown that Cascade K-means algorithm achieve 87,86% of OA and Kappa 0,56, and Naive Bayes with 88,59% of OA and Kappa 0,58.

Palavras-Chave - *Naive Bayes, Cascade K-Means, Google Earth Engine.*

1. INTRODUÇÃO

As imagens de satélites são uma alternativa viável e de baixo custo para o estudo do uso e ocupação do solo, auxiliando nas análises dos impactos das ações antrópicas em bacias

hidrográficas [1], monitorar o crescimento e desenvolvimento urbano [2], como também no segmento agrícola, abrangendo o levantamento ou identificação de solos, cobertura da vegetação, estimativa de biomassa, produtividade, estresse hídrico, vigor das plantas e a avaliação do estado fenológico das culturas [3]. A extração de informações de imagens pode ser facilitada por técnicas de classificação de imagens, as quais reduzem o tempo necessário e a subjetividade da interpretação humana. Classificadores subdividem-se em supervisionados e não-supervisionados, em que estes extraem, identificam e classificam os pixels com base na amostra de treinamento e pela análise de características semelhantes entre si [4] ou de forma automática sem um conhecimento prévio dos alvos, respectivamente.

Um dos algoritmos de classificação não-supervisionada para sensoriamento remoto, é o *K-means*. Este algoritmo agrupa os dados em K dimensões do espaço, onde K refere-se ao número de bandas espectrais e/ou imagens envolvidas [5]. Em contrapartida, um classificador supervisionado comumente utilizado é o algoritmo *Naive Bayes*, que considera todos os atributos independentes entre si dado o valor da variável da classe [6].

A plataforma Google Earth Engine (GEE) vem revolucionando o processamento de dados geoespaciais por proporcionar facilidade, extensa documentação e agilidade de processamento. Logo, o presente trabalho teve como intuito empregar técnicas de classificação supervisionada e não-supervisionada para construir o mapa temático de uso e ocupação do município de Tupãssi - PR, para o ano-safra 2016/2017 com imagens do sensor Landsat-8/OLI pela plataforma GEE, visando reduzir o tempo atrelado ao processamento, despontando como uma ferramenta ao suporte, planejamento e monitoramento do uso e cobertura do solo.

2. MATERIAL E MÉTODOS

O município de Tupãssi, localizado na mesorregião do oeste do estado do Paraná, possui 310 km² de área, tendo como principais atividades: cultivo de soja, milho e feijão, avicultura, suinocultura e bovinocultura [7]. O clima do município é classificado como Cfa, ou seja, possui um clima subtropical úmido mesotérmico, com verões quentes e geadas

pouco frequentes, com tendência de concentração de chuvas nos meses de verão, sem estação seca. A média das temperaturas dos meses mais quentes é superior a 22° C e dos meses mais frios é inferior a 18° C [8].

Cenas dos satélites Landsat-8/OLI e Sentinel-2 foram empregadas para a classificação manual do município (ou seja, vetorização dos alvos sob estudo), e do satélite Landsat-8/OLI, produto TOA, para as classificações não-supervisionada e supervisionada. Os alvos mais recorrentes em Tupãssi foram: soja (17 áreas), milho (6 áreas), vegetação (nativa e reflorestamento, 15 áreas), água (açudes, 7 áreas) e edificações (rurais e urbanas, 7 áreas), formando assim cinco classes de interesse.

Tendo em vista a necessidade de identificar culturas agrícolas, utilizou-se cenas de 01/11/16 a 01/03/17, período este que abrange o ciclo de desenvolvimento das culturas sob estudo [9]. Outro critério foi a porcentagem de nuvens das cenas, que deveriam ser inferiores a 15%. Totalizando assim, três cenas para o satélite Landsat-8/OLI (24/11/16, 11/01/17 e 27/01/17) e duas cenas do sensor Sentinel-2 (15/12/16, 02/01/17). Na plataforma GEE executou-se a remoção de nuvens e sombra de nuvens das cenas do satélite Landsat-8/OLI pela linguagem *javascript*. Para as cenas de ambos os satélites foi calculado a mediana das imagens, e desta se prosseguiu com a classificação.

A identificação visual dos alvos se deu pela falsa-cor RGB-8-11-4 para Sentinel-2 e RGB-5-6-4 para o Landsat-8. Ambas as imagens foram aplicadas o filtro *percent clip* (mínimo: 0,5 e máximo:1,0) para realçar o contraste das imagens falsa-cor (Figura 1) no software ArcGis 10.1. Com isto, os alvos das cinco classes foram vetorizados e transformados para raster, como também a geração dos mapas temáticos.

Para a classificação, foram usadas seis bandas do satélite Landsat-8: 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), como também os índices de vegetação *Normalized Difference Vegetation Index* (NDVI, Equação 1), *Soil-Adjusted Vegetation Index* (SAVI, Equação 2), *Enhanced Vegetation Index* (EVI 2, Equação 3) e *Normalized Soil Difference Index* (NSDI, Equação 4). O emprego de índices distintos se dá pelo ganho na detecção dos alvos que cada um proporciona.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

$$SAVI = \frac{NIR - RED}{NIR + RED + L} * (1 + L) \quad (2)$$

$$EVI2 = \frac{NIR - RED}{(NIR + 2,4 * RED + 1)} \quad (3)$$

$$NSDI = \frac{SWIR2 - BLUE}{SWIR2 + BLUE} \quad (4)$$

Em que: NIR: refletância no infravermelho próximo, RED: refletância na região do vermelho, L: fator de correção do solo, SWIR: refletância do infravermelho próximo, BLUE: refletância da banda azul.

O NDVI, possibilita uma diferenciação entre vegetação e demais alvos, sendo amplamente utilizado para monitorar o uso e ocupação do solo, detecção de épocas secas, como também para o monitoramento de culturas agrícolas [10]. Todavia, as condições de fundo do solo exercem considerável influência sobre as leituras de refletâncias de culturas agrícolas dos satélites. Para contornar isto, o índice SAVI, além de considerar a refletância da vegetação, possui um fator de correção para reduzir a influência do solo no cálculo.

O índice EVI-2, proposto por Jiang et al. [11], é derivado do índice EVI, e não utiliza a banda azul para reduzir ruídos e a influência de aerossóis atmosféricos [12]. Logo, o EVI-2 possui grande sensibilidade em áreas com alto valor de biomassa por realçar a maior absorção da radiação eletromagnética pela clorofila [13]. O índice NSDI, proposto por Rogers e Kearney [14] e adaptado por Deng et al. [15] foi utilizado com o intuito de destacar as áreas construídas e o solo exposto dos demais alvos.

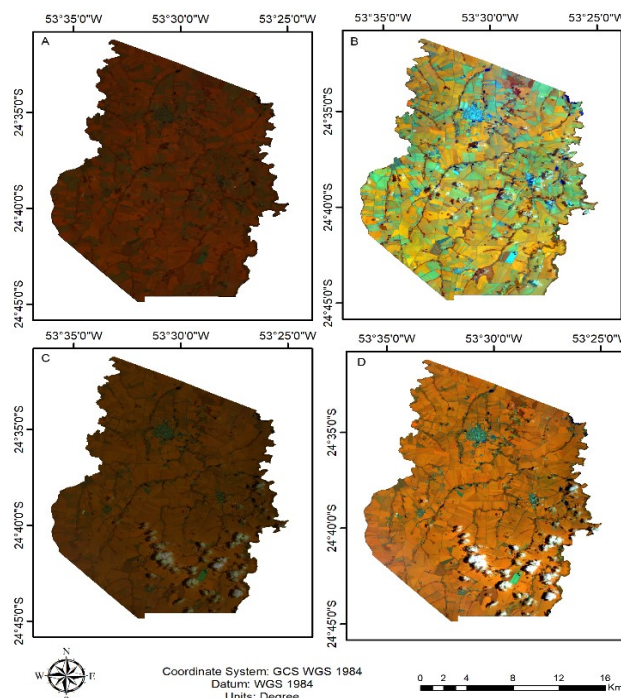


Figura 1. Composição falsa cor e aplicação do filtro realce em: (A) RGB 564 Landsat 8/OLI, (B) RGB 564 Landsat 8/OLI com realce, (C) RGB 8114 Sentinel-2, (D) RGB 8114 Sentinel-2 com realce.

A classificação não-supervisionada foi executada pelo classificador *Cascade K-Means*, o qual segregou os alvos presentes na imagem em 15 classes distintas, as quais foram avaliadas posteriormente e distribuídas para as cinco classes pré-definidas. Este algoritmo é uma modificação do algoritmo *K-means* que cria diversas partições de pequenos a grandes grupos, formando assim uma cascata [16]. A classificação supervisionada dos alvos do município foi executada pelo algoritmo *Naive Bayes* para variáveis contínuas [17]. Todo o processamento foi executado na plataforma GEE. A acurácia dos mapeamentos foi avaliada

pelo índice Kappa [18] e pela a Exatidão Global (EG), computadas pixel a pixel.

3. RESULTADOS E DISCUSSÃO

A mesorregião do Oeste do Paraná caracteriza-se por apresentar índices de precipitação bem distribuídos ao longo do ano, o que acarreta na probabilidade de existência de chuvas com nuvens ao longo do ciclo da cultura. A presença de nuvens numa imagem modifica a reflectância espectral de um alvo, dificultando a detecção de feições, alterando também a contabilização da área na classificação da imagem [19]. Entretanto, o processamento utilizado para a remoção de nuvens mostrou-se efetivo, visto que boa parte das áreas que possuíam nuvens foram classificadas corretamente pelos algoritmos utilizados (Figura 1B e Figura 2).

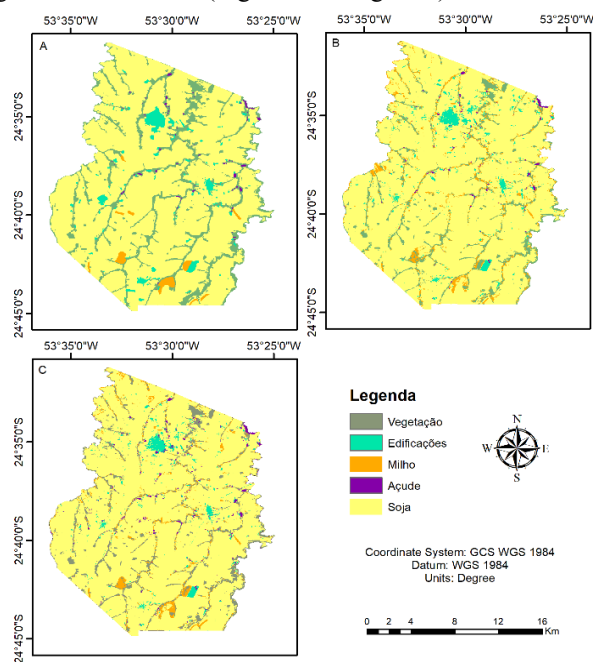


Figura 2. Mapa temático vetorizado (A), classificação não-supervisionada *Cascade K-Means* (B), classificação supervisionada *Naive Bayes* (C).

Ambos os algoritmos (*Cascade K-Means* e *Naive Bayes*) apresentaram confusão na detecção dos alvos milho e vegetação, ocasionando em erros de omissão e inclusão entre as classes. Analisando as áreas obtidas (Tabela 1), tem-se que ambas as técnicas foram coerentes na detecção da área de soja (85,68% e 86,77% para a *Cascade K-Means* e *Naive Bayes*, respectivamente) em comparação a classificação manual (82,59%). Todavia, as classes de milho e vegetação para ambas as classificações apresentaram as maiores discrepâncias quando comparadas a classificação manual.

Avaliando a área obtida das classes açude e vegetação pelos algoritmos e pela vetorização (Tabela 1), estas demonstram-se próximas, ilustrando assim a capacidade de ambos em diferenciarem alvos agrícolas e não-agrícolas.

Analisando a diferença percentual entre as técnicas de classificação utilizadas (Tabela 1) quando comparado com a

vetorização, tem-se que o algoritmo *Cascade K-Means* obteve um menor erro percentual para as classes cidade, açude e soja (3,9%, 25,5%, 3,7%, respectivamente). Para as classes vegetação e milho, o algoritmo *Naive Bayes* gerou menor erro, embora muito elevado (45,6% e 178,6%).

Tabela 1. Área e percentual de área das classes obtidas através da vetorização, e diferenças percentuais entre as áreas das classificações automática e supervisionada

Classe	Vetorização		<i>Cascade K-Means</i>			<i>Continuos Naive Bayes</i>		
	Área (ha)	% ¹	Área (ha)	% ¹	% ²	Área (ha)	% ¹	% ²
Vegetação	4.551	13,2	2.358	6,9	48,2	2.477	7,2	45,6
Edificações	944	2,7	908	2,6	3,9	721	2,1	23,6
Milho	375	1,1	1.507	4,4	302,0	1.045	3,0	178,6
Açude	127	0,4	159	0,5	25,5	315	0,9	147,8
Soja	28.440	82,6	29.505	85,7	3,7	29.880	86,8	5,1

%¹: Porcentagem de área de cada classe pré-definida obtida nas diferentes vetorizações.

%²: Diferenças percentuais entre as classificações automática e supervisionada, tendo como referência a classificação visual.

Estes fatos podem ser observados nas Tabelas 2 e 3, que representam as matrizes de confusão obtidas na comparação dos mapeamentos com a referência (a vetorização manual). Também constam nas tabelas as Acurácias de Produtor (AP), que aponta a probabilidade de um pixel ter sido classificado corretamente e a Acurácia de e Usuário (AU) que aponta a probabilidade de um pixel classificado corresponder a classe de referência.

Tabela 2. Matriz de confusão das classes vetorizadas manualmente e pelo classificador *Cascade Kmeans*

		Referência					AU
		Floresta	Cidade	Milho	Açude	Soja	
Mapeamento	Floresta	22052	178	696	55	3022	0,85
	Cidade	1072	5609	60	127	3214	0,56
	Milho	9765	430	1982	100	4381	0,12
	Açude	619	121	0	967	61	0,55
	Soja	16798	4139	1428	148	305268	0,93
	Total	50306	10477	4166	1397	315946	
AP		0,44	0,54	0,48	0,69	0,97	

Tabela 3. Matriz de confusão das classes vetorizadas manualmente e pelo classificador *Continuos Naive Bayes*

		Referência					AU
		Floresta	Cidade	Milho	Açude	Soja	
Mapeamento	Floresta	23174	228	411	39	3456	0,85
	Cidade	804	5059	70	51	2024	0,63
	Milho	5421	111	2679	3	3372	0,23
	Açude	1450	506	0	1106	423	0,32
	Soja	19457	4573	1006	198	306671	0,92
	Total	50306	10477	4166	1397	315946	
AP		0,46	0,48	0,64	0,79	0,97	

A classe milho apresenta como a maior fonte de erro, confundindo principalmente com as classes floresta e soja. Isto se deve principalmente a modificação da reflectância espectral dos alvos causada pelas nuvens, uma vez que o processo de remoção de nuvens, apesar de eficiente, não foi completo, não retirando nuvens finas. Outro fator que levou a baixa acurácia dos resultados foi a mistura espectral do

pixel, em que em áreas de borda de floresta ou talhão de soja foi classificado como milho.

As demais inconsistências foram, em sua maioria, provenientes da diferente resolução da vetorização e mapeamento. A vetorização manual, realizada a partir de imagens com resolução espacial de 10 metros (Sentinel-2) e, quando preciso, 30 metros (Landsat-8), foi muito mais apurada que a classificação obtida exclusivamente com imagens Landsat-8. Desta forma, apesar dos resultados serem muito próximos quando comparados visualmente, apresentaram diferenças na delimitação das áreas.

Tais problemas refletiram na EG e Kappa dos mapeamentos, em que o classificador *Cascade K-means* obteve EG de 87,86% e Kappa 0,56 e o classificador *Naive Bayes* um EG de 88,59% e Kappa 0,58.

4. CONCLUSÕES

A metodologia empregada demonstrou-se eficaz no mapeamento de áreas agrícolas. O município de Tupãssí tem predomínio agrícola, com mais de 80% de área com cultivo de soja. Ambas as técnicas apresentaram resultados semelhantes, com baixas acurácias nas classes vegetação e milho. A elevada confusão entre estas classes indica a necessidade de um maior estudo sobre técnicas de aprendizado de máquina para a busca de algoritmos mais eficazes na identificação dos alvos. Neste sentido, o presente estudo atende ao propósito de gerar planos de informações de maneira ágil a todos os membros da cadeia agropecuária.

5. AGRADECIMENTOS

A UNIOESTE - Campus Cascavel, a Pós-Graduação em Engenharia Agrícola (PGEAGRI) e ao Núcleo de Pesquisa GeoScience pela infraestrutura e apoio técnico-científico. A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), ao CNPq e a Fundação Araucária (FA) pelo apoio financeiro.

6. REFERÊNCIAS

- [1] Coelho, V. H. R., Montenegro, S., Almeida, C. N. “Dinâmica do uso e ocupação do solo em uma bacia hidrográfica do semiárido brasileiro”. *Revista Brasileira de Engenharia Agrícola e Ambiental*, volume (18), número (1), páginas (64–72), 2014.
- [2] Júnior, L. N. “Urbanização e cidade dispersa: implicações da produção do espaço urbano no Brasil, em Moçambique e na Austrália”. *GeoUSP*, volume (21), número (2), páginas (550–569), 2017.
- [3] Bernardi, A. C. de C., Naime, J. M., Resende, A. V., Bassoi, L. H., Inamasu, R. Y. “Agricultura de precisão: resultados de um novo olhar”. Brasília, DF: Embrapa, capítulo (4), páginas (58-73), 2014.
- [4] Meneses, P. R., Almeida, T., “Introdução ao processamento de imagens de sensoriamento remoto”, UNB, Brasília, páginas (266), 2012.
- [5] Schowengerdt, R. A. “Techniques for image processing and classification in remote sensing”. Academic Press, New York, páginas (249), 1983.
- [6] Da Silva, C. C., Oliveira, S. R. de M., Adami, S. F., Crivelenti, R. C., Coelho, R. M. “Diferentes classificadores na predição de classes de solos em mapeamento digital”. In *Anais: XV Simpósio Brasileiro de Sensoriamento Remoto – SBSR*. Curitiba, PR, 2011
- [7] IBGE - Cidades. Disponível em: <<https://cidades.ibge.gov.br/brasil/pr/Tupãssí/panorama>>. Acesso em: 28 de junho de 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] SEAB, S. de E. da A. e do A. SOJA - Análise da Conjuntura. p. 1–6, 2018. Disponível em: <http://www.agricultura.pr.gov.br/arquivos/File/deral/Prognosticos/2018/Soja_2017_18.pdf>.
- [10] Gandhi, G. M., Parthiban, S., Thummalu, N., Christy, A. “NDVI: Vegetation Change Detection Using Remote Sensing and Gis - A Case Study of Vellore District”. *Procedia Computer Science*, volume (57), páginas (1199–1210), 2015.
- [11] Jiang, Z., Huete, A. R., Didan, K., Miura, T. “Development of a two-band enhanced vegetation index without a blue band”. *Remote Sensing of Environment*, volume (112), número (10), páginas (3833–3845), 2008.
- [12] Albuquerque, C. K., Cruz, C. B. M., Barros, R. S. “Comparação do comportamento do NDVI e do EVI-2, um novo índice de vegetação”. In *Anais III Simpósio Brasileiro de Ciências Geodésicas e Tecnologia de Geoinformação*. Recife, PE, páginas (01-04), 2010.
- [13] Paula, S. C. de, Vedovato, L. B., Silva, G. B. S. da, Vicente, L. E. “Comparação entre os índices espectrais EVI e EVI2 para discriminação de pastagens plantadas e naturais”. In *Anais: VII Congresso Interinstitucional de Iniciação Científica - CIIC 2013*, páginas (1–8), 2013.
- [14] Rogers, A. S., Kearney, M. S. “Reducing signature variability in unmixed coastal marsh Thematic Mapper scenes using spectral indices”. *International Journal of Remote Sensing*, volume (25), número (12), páginas (2317–2335), 2004.
- [15] Deng, Y., Wu, C., Li, M., Chen, R. RNSDI: “A ratio normalized difference soil index for remote sensing of urban/suburban environments”. *International Journal of Applied Earth Observation and Geoinformation*, volume (39), páginas (40–48), 2015.
- [16] Martins, A.C., Marques, M. J., Costa, P.D. “Estudo comparativo de três algoritmos de machine learning na classificação de dados electrocardiográficos”. Dissertação (Mestrado) Universidade do Porto, 2009.
- [17] McCulloch, D. R., Lawry, J., Rico-Ramírez, M. A., Cluckie, I. D. “Detecting Bright Band using AI techniques in radar hydrology”. *Remote Sensing for Environmental Monitoring and Change Detection*, páginas (37–46), 2007.
- [18] Congalton, R. G. “A review of assessing the accuracy of classifications of remotely sensed data”. *Remote Sensing of Environment*, volume (49) número (12), páginas (1671-1678), 1991.
- [19] Silva, M. A. O., Liporace, F. dos S. “Automatic detection of cloud and cloud shadow in remote sensing images”. *Boletim de Ciências Geodésicas*, volume (22), número (2), páginas (369–388), 2016.