

## MEDIANA E DESVIO PADRÃO DO PERFIL ESPECTRO-TEMPORAL DE NDVI COMO PARAMETROS DE CLASSIFICAÇÃO

Willyan Ronaldo Becker<sup>1</sup>, Ivã Luis Caon<sup>2</sup>, Carlos Eduardo Vizzotto Cattani<sup>1</sup>, Erivelto Mercante<sup>3</sup>, Jerry Adriani Johann<sup>3</sup>, Diandra Ganascini<sup>1</sup>, Victor Hugo Rohden Prudente<sup>4</sup>

<sup>1</sup>Universidade Estadual do Oeste do Paraná, Programa de Pós-Graduação em Engenharia Agrícola, Cascavel, Paraná, Brasil, {willyan.becker@unioeste.br; carlos.cattani@unioeste.br; diandra.ganascini@unioeste.br}; <sup>2</sup>Universidade Estadual do Oeste do Paraná, Graduando em Engenharia Agrícola, Cascavel, Paraná, Brasil, {iva.caon@unioeste.br}; <sup>3</sup>Universidade Estadual do Oeste do Paraná, Docente no Programa de Pós-Graduação em Engenharia Agrícola, Cascavel, Paraná, Brasil, {erivelto.mercante@unioeste.br; jerry.johann@unioeste.br}; <sup>4</sup>Instituto Nacional de Pesquisas Espaciais, São José dos Campos, São Paulo, Brasil, {victor.prudente@inpe.br}

### RESUMO

Este trabalho teve como objetivo a geração de um mapeamento de uso e cobertura da terra para a bacia hidrográfica do rio São Francisco Verdadeiro (SFV), a partir da plataforma de processamento geoespacial *Google Earth Engine* (GEE), utilizando como parâmetro a mediana e o desvio padrão da série temporal do índice de vegetação por diferença normalizada (NDVI). Os resultados mostram que a maior parte do território da bacia é agricultura (60,7%), seguida de mata (20,3%), pastagem (18,1%) e corpos hídricos (0,9%). e. A exatidão global da classificação foi de 97%, o índice Kappa de 0,96 e índice Tau de 0,96. Conclui-se que as métricas utilizadas demonstram ser bons parâmetros para classificação por Sensoriamento Remoto (SR) e que a plataforma GEE é uma excelente ferramenta para execução de processamentos geoespaciais, devido a sua alta capacidade de processamento.

**Palavras-chave** — Bacia hidrográfica, Google Earth Engine, Random Forest, Uso e cobertura da terra.

### ABSTRACT

*The objective of this work was the generation of a land use and land cover mapping for the São Francisco Verdadeiro river basin, on the geospatial processing platform Google Earth Engine (GEE), using as a parameter the mean and the standard deviation of the time series of the Normalized Difference Vegetation Index. The results show that most of the territory of the basin is agriculture (60.7%), followed by forest (20.3%), pasture (18.1%) and water bodies (0.9%). The overall accuracy of the classification was 97%, the Kappa index was 0.96 and the Tau index was 0.96, whose results were considered excellent. It is concluded that the metrics used are good parameters for classification by remote sensing and that the GEE platform is an excellent tool for execution of geospatial processes due to its high processing capacity.*

**Key words** — Watershed, Google Earth Engine, Random Forest, Land use and land cover.

### 1. INTRODUÇÃO

A forma de uso, tipos de solo e relevo, a vegetação local existente, o desmatamento e a presença de cidades exercem grande pressão sobre os recursos naturais que compõem uma bacia hidrográfica. Todas as atividades realizadas na bacia desenvolvida por

indústrias, propriedades rurais e cidades refletem na qualidade da água do rio, desde suas nascentes até a sua foz [1].

A importância da análise do uso e da cobertura da terra, principalmente em bacias hidrográficas, justifica-se na necessidade de identificação de fontes ou potenciais fontes de alterações do ambiente, advindas das atividades humanas [2].

Uma das formas de se analisar o uso e cobertura da terra é por meio do Sensoriamento Remoto (SR). O SR é a ciência de obter informações sobre um objeto, área ou fenômeno através da análise de dados adquiridos por um sensor que não está em contato com o objeto, área ou fenômeno sob investigação. Assim, o objetivo de qualquer tecnologia de SR é fornecer observação de algum parâmetro físico em algum tempo ou período de tempo [3].

Os alvos terrestres, quando estudados em SR, possuem uma assinatura espectral definida, que pode ser denominando como padrão da reflectância. A partir de índices de vegetação, juntamente com estudo em múltiplas datas, podemos traçar o padrão espectro-temporal dos alvos, contribuindo na identificação dos mesmos. O índice de NDVI (*Normalized Difference Vegetation Index*) envolve a razão da diferença e da soma entre as bandas do infravermelho próximo e do vermelho [4]. Este índice contribui para o destaque da vegetação em relação aos demais alvos presentes no solo, bem como a redução dos possíveis erros dos instrumentos e de calibração, por meio da razão matemática que normalizado os dados [5].

O desvio padrão é uma métrica estatística que indica a dispersão dos dados em torno da média. Um baixo desvio padrão indica que os dados tendem a estar próximos da média ou do valor esperado. Já um alto desvio padrão indica que os dados estão espalhados por uma ampla gama de valores [6]. A métrica de desvio padrão de imagens de satélite é uma variável computacionalmente custosa, principalmente se envolver uma grande quantidade de imagens. Entretanto, a computação em nuvem vem proporcionando capacidades de processamento de dados cada vez maiores, ampliando o leque de possibilidades em SR [7].

A plataforma de processamento geoespacial baseada em nuvem *Google Earth Engine* (GEE) trata-se de plataforma em nuvem com objetivo de fazer análise de dados ambientais em grande escala e com recursos computacionais amplos, utilizando os servidores Google para processamento. A análise e processamento de dados por meio do GEE facilita uma abordagem interativa para o desenvolvimento de algoritmos e um

ciclo rápido de testes e melhorias para o processamento em larga escala [7].

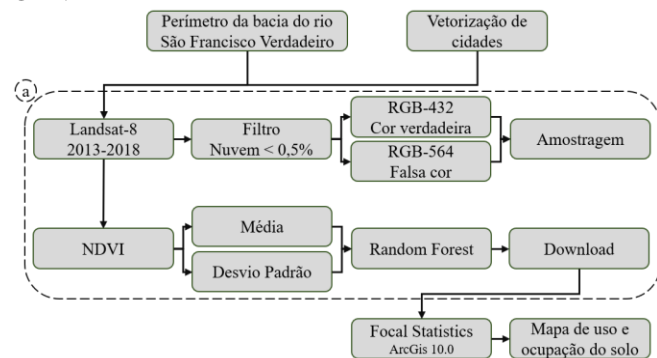
Assim, o objetivo deste trabalho foi realizar o mapeamento do uso e cobertura da terra para a bacia do rio São Francisco Verdadeiro (SFV) com a utilização da mediana e do desvio padrão do índice de vegetação NDVI entre os anos de 2013 a 2018 como parâmetros de diferenciação dos alvos, dentro do projeto “Desenvolvimento de metodologia para análise de uso e cobertura da terra com imagens de satélite”, CP 02/2014 Apoio a Concessão de Bolsas FA - FPTI do PTI.

## 2. MATERIAIS E MÉTODOS

A área de estudo compreendeu a bacia hidrográfica do rio SFV, localizado na mesorregião Oeste do Estado do Paraná. A bacia SFV está presente na área territorial de 11 municípios, possui uma área total de 2.219,11 km<sup>2</sup>, caracterizando-a como a maior bacia em extensão dentro da Bacia Hidrográfica Paraná 3, com 152 quilômetros, com sua nascente na cidade de Cascavel e sua foz no município de Entre Rios do Oeste, onde o rio São Francisco Verdadeiro desagua no rio Paraná, já nas dependências do reservatório da Usina Hidrelétrica de Itaipu [1]. Nesta região há predominância da agricultura de ciclo anual, principalmente com monoculturas (soja, milho e trigo).

A bacia está inclusa no programa HELP (Hidrologia para Ambiente, Vida e Política – do inglês *Hydrology for the Environment, Life and Policy*), uma iniciativa da UNESCO que fomenta o gerenciamento de bacias hidrográficas em todo o mundo de acordo com critérios sustentáveis [8,9].

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



**Figura 1. Fluxograma de trabalho de mapeamento de uso e cobertura da terra para a bacia do rio SFV. Em (a) as etapas realizadas na plataforma Google Earth Engine.**

O clima na bacia do rio SFV, segundo a classificação de Köppen, se enquadra no tipo Cfa [10]. Quanto ao relevo, compreende o Terceiro planalto paranaense, que recobre a maior porção do Estado do Paraná [11] 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 [12].

O perímetro da bacia foi cedido por Cattani et al. [13], os quais realizaram a caracterização morfométrica. De início,

foram vetorizadas todas as cidades, vilas e grandes áreas construídas no interior da área de abrangência da bacia, com o *software Google Earth Pro*, para que estas não viessem a ser erroneamente classificadas e contabilizadas em uma classe de interesse.

Utilizou-se toda a série de imagens ópticas sensor *Operational Land Imager (OLI)*, a bordo do satélite Landsat-8. As imagens OLI possuem resolução espacial de 30 metros e temporal de 16 dias. A execução se deu na plataforma web de processamento de dados geoespaciais GEE. Foi definido o percentual máximo de nuvem de 0,5% nas imagens, para que estas não influenciassem nas métricas a ser calculadas. Ao todo, foram utilizadas 41 imagens, de abril de 2013 a maio de 2018. A seleção dos dados de treinamento realizou-se mediante composição RGB-432 (cor verdadeira) e RGB-564 (falsa cor), que proporcionou a identificação visual dos alvos em função da sua interação com a radiação eletromagnética. Foram coletadas 50 amostras de Mata (nativa e reflorestamento), 35 de Corpos Hídricos, 20 de pastagem e 60 de Agricultura, sendo tais classes definidas para coincidir com as apresentadas pelo Centro Internacional de Hidroinformática (CIH) [11].

Calculou-se, então, o índice de vegetação NDVI para cada uma das 41 imagens, gerando perfis de NDVI. Os perfis de NDVI de classes de uso da terra como a agricultura representam ciclos, uma vez que os alvos são altamente variáveis. Neste sentido, esta abordagem tem a ideia de se beneficiar desta variabilidade para as etapas de classificação.

Para reduzir o conjunto de dados, foi calculado a mediana e o desvio padrão da série de NDVI, sendo estas as duas únicas variáveis empregadas para o processo de classificação. O desvio padrão é uma medida estatística que indica a dispersão dos dados em torno da média, sendo empregada neste estudo como uma variável para diferenciação dos alvos com comportamento espectro-temporal cíclico. A variável mediana situa a posição do perfil no gráfico espectro-temporal do índice de vegetação, contribuindo para diferenciar alvos com valores de desvio padrão semelhantes.

O mapeamento do uso e cobertura da terra foi realizado por meio do uso do algoritmo de classificação *Random Forest (RF)*, o qual se destaca devido a sua robustez e capacidade de lidar com alto número de variáveis [14]. O RF cria várias árvores de decisão, treinadas por meio da seleção aleatória de dois terços dos dados de treinamento, enquanto o terço restante é empregado na validação cruzada da árvore gerada [14]. Foi definido o total de 100 árvores para a classificar a área de estudo [15]. Após o processo de classificação, o resultado foi enviado ao *Google Drive* e posteriormente realizado o download.

Após a classificação e download do resultado, foi utilizado o *software ArcGIS 10.0* para a aplicação do filtro *Focal Statistics* a fim de retirar os ruídos ocasionados pelas possíveis confusões na classificação.

A etapa de avaliação da qualidade da classificação consistiu no sorteio de 100 pontos aleatórios em cada classe

estudada e, por meio de inspeção visual em imagens de alta resolução do *Google Earth Pro* foram confirmados ou não a real classificação dos pontos. 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) [16,17].

Conforme Congalton [16], a acurácia de produtor indica a probabilidade de um pixel de avaliação ter sido classificado corretamente. A acurácia de usuário indica a probabilidade de um pixel classificado corresponder a classe avaliada. A exatidão global avalia o quanto o mapeamento obteve de acerto. Os índices Kappa e Tau avaliam a proporção de classificações corretas, desconsiderando o acerto ao acaso [16,17].

### 3. RESULTADOS E DISCUSSÃO

A Figura 2 mostra o resultado final da classificação que contou com os parâmetros de mediana e desvio padrão da série de índice de vegetação NDVI para diferenciação dos alvos na bacia do rio SFV.

Vemos que a classe preponderante é a de Agricultura, com 60,7% da área da bacia, seguida da área de Mata (20,3%), Pastagem (18,1%) e Água (0,9%). Tais resultados são proporcionalmente similares aos apresentados no portal do CIH [11], com dados de 2006, no qual 72% da área da bacia, já desconsiderando a área urbana, era de uso agrícola. A mata é apresentada como 17% da área, pastagem 10% e água 2% da área [11].

Por análise visual da Figura 2, bem como pela Tabela 1, podemos concluir que as maiores áreas de pastagem estão presentes nos municípios de Marechal Cândido Rondon, São José das Palmeiras e Ouro Verde do Oeste, região central/sudoeste da bacia. Já as áreas de mata e agricultura estão bem dispersas pelos demais municípios, sendo Toledo do município com mais área nestas classes, com 28,7% de

Mata e 34,4% de agricultura. A maior parte dos Corpos Hidricos (classe Água) encontra-se nos municípios de Pato Bragado e Entre Rios do Oeste, ambos presentes na foz do rio São Francisco Verdadeiro, onde este se encontra com o reservatório da usina de Itaipu. Também apresenta com maior percentual de água o município de Toledo, sendo nele presente a Pequena Central Hidreletrica São Francisco, com 2 unidades geradoras e potência total de 14,0 MW [18].

Tabela 1. Percentual de área de cada classe em cada município presente na bacia hidrográfica do rio São Francisco Verdadeiro.

Município	Mata	Água	Pastagem	Agricultura
Cascavel	6,6%	0,2%	6,1%	15,2%
Entre Rios do Oeste	4,2%	43,3%	2,5%	3,8%
Marechal Cândido Rondon	20,7%	3,2%	16,3%	14,3%
Ouro Verde do Oeste	13,1%	0,6%	21,0%	11,9%
Pato Bragado	4,0%	50,0%	2,0%	3,5%
Quatro Pontes	1,4%	0,0%	1,4%	1,4%
Santa Helena	0,8%	0,0%	1,8%	0,2%
Santa Tereza do Oeste	5,3%	0,0%	4,2%	8,1%
São José das Palmeiras	6,4%	0,0%	14,8%	1,2%
São Pedro do Iguacu	8,8%	0,2%	9,8%	6,0%
Toledo	28,7%	2,5%	20,1%	34,4%
<b>Total</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>

Nota: Percentual referente ao total da bacia hidrográfica.

Conforme avaliação dos 100 pontos aleatórios criados para cada classe, vistoriados em imagens de alta *Google Earth Pro*, foi criada a matriz de confusão (Tabela 2) e obtidas as estatísticas de acurácia.

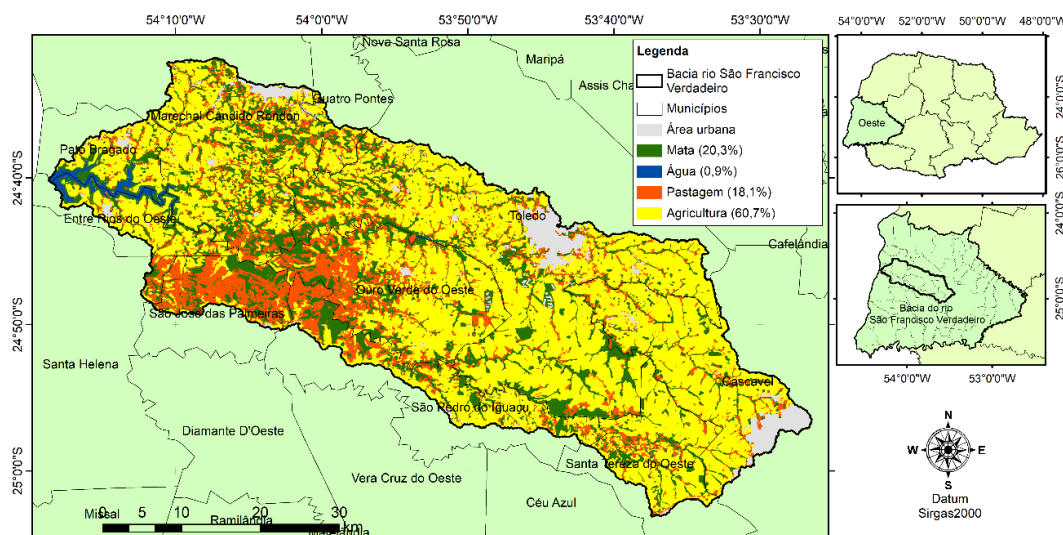


Figura 2. Mapeamento do uso e cobertura da terra da bacia do rio SFV com parâmetros de classificação mediana e desvio padrão de NDVI.



A classe pastagem representa a maior fonte de erro, confundindo principalmente com a classe de agricultura. Isto se deve a sazonalidade das pastagens, que influenciam na série de NDVI e consequentemente ocasionam pequenas confusões com a agricultura.

Tabela 2. Matriz de confusão das classes mapeadas e verificadas em imagens de alta resolução Google Earth Pro

		Referência				Total	AU
		Mata	Água	Pastagem	Agricultura		
Mapeamento	Mata	97		2	1	100	<b>0,97</b>
	Água		100			100	<b>1</b>
	Pastagem	2		95	3	100	<b>0,95</b>
	Agricultura	1	1	2	96	100	<b>0,99</b>
	Total	100	101	99	100	400	
AP		<b>0,97</b>	<b>0,99</b>	<b>0,96</b>	<b>0,96</b>		

## 5. CONCLUSÕES

O mapeamento de uso e cobertura da terra empregando métricas estatísticas de mediana e desvio padrão de séries temporais de imagens de satélite se mostrou uma ótima opção, alcançando excelentes resultados.

A plataforma GEE possibilitou a execução do mapeamento de forma rápida e fácil. A interatividade da plataforma e o rápido processamento foram cruciais para a etapa de aprendizado e da aplicação, mesmo com a extensa lista de imagens utilizadas e todos os processamentos realizados (filtro de seleção, recorte, mediana, desvio padrão, treinamento e classificação).

## 6. AGRADECIMENTOS

Os autores agradecem a Universidade Estadual do Oeste do Paraná (UNIOESTE), a Pós-Graduação em Engenharia Agrícola (PGEAGRI), ao Parque Tecnológico Itaipu (PTI) e ao Laboratório de Topografia e Sensoriamento Remoto (GeoLab) pelo apoio técnico-científico e a Fundação Araucária (FA), a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Código de Financiamento 001 e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) 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] Santos, L. T. S. de O.; Jesus, T. B. DE; Nolasco, M. C. "Influência do Uso e Ocupação do Solo na Qualidade das Águas Superficiais Do Rio Subaé, Bahia" *Geographia Opportuno Tempore*, volume (1), número (1), páginas (68–79), 2014.
- [3] Jensen, J.R., "Sensoriamento Remoto do Ambiente: uma perspectiva em Recursos Terrestres", Tradução de J. C. N. Epiphânio. São José dos Campos, SP: Parênteses, páginas (598), 2009. Tradução de: *Remote Sensing of the Environment: an Earth Resource Perspective* (Prentice Hall Series in Geographic Information Science).
- [4] Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. "Monitoring vegetation systems in the Great Plains with ERTS". *Third Earth Resources Technology Satellite-1 Symposium*, 309-327, Washington D.C., 1973.

[5] Didan, K.; Munoz, A.B.; Solano, R.; Huete, A. "MODIS Vegetation Index User's Guide (MOD13 Series)". The University of Arizona, 2015.

[6] PORTAL ACTION. "Medidas de dispersão". Disponível em: <<http://www.portalaction.com.br/estatistica-basica/22-medidas-de-dispersao>>. Acesso em: 4 out. 2018.

[7] 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.

[8] UNESCO, United Nations Educational Scientific and Cultural Organization; KHAN, S. "HELP - Hydrology for the Environment, Life and Policy". 2010. Disponível em: <<http://unesdoc.unesco.org/images/0021/002145/214516E.pdf>>. Acesso em: 14 mar. 2018

[9] UNESCO, United Nations Educational Scientific and Cultural Organization. "Policy (HELP)". Disponível em: <<https://en.unesco.org/themes/water-security/hydrology/programm-es/help>>. Acesso em: 14 mar. 2018.

[10] 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.

[11] CIH, C. I. de H. "Bacia rio São Francisco Verdadeiro". Disponível em: <<http://saofrancisco.hidroinformatica.org.br/>>. Acesso em: 5 set. 2018.

[12] 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.

[13] Cattani, C. E. V.; Mercante, E.; Oldoni, L. V.; Almeida, L.; Mendes, I. de S.; Scariotto, M. C. "Uso de Técnicas de Geoprocessamento para Caracterização Morfométrica da Bacia do São Francisco Verdadeiro". In: Sione, W. F.; Mayer, F. M. V.; Antes, M. E.; Serafini, M. C. (Ed.). XVII Sociedad Latinoamericana de Especialistas Percepción Remota y Sistemas de Información Espacial – SELPER. 13 páginas. EdUnLu, 2016.

[14] Breiman, L. "Random forests". *Machine learning*, volume (45), número (1), páginas (5–32), 2001.

[15] Onojeghuo, A. O.; Blackburn, G. A.; Wang, Q.; Atkinson, P. M.; Kindred, D.; Miao, Y. "Mapping paddy rice fields by applying machine learning algorithms to multi-temporal sentinel-1A and landsat data". *International Journal of Remote Sensing*, volume (39), número (4), páginas (1042–1067), 2018.

[16] Congalton, R.G. "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sens. Environ.*, vol. 37, no. 1, pp. 35–46, 1991.

[17] Souza, C. H. W.; Mercante, E.; Prudente, V. H. R.; Justina, D. D. D. "Methods of performance evaluation for the supervised classification of satellite imagery in determining land cover classes". *Ciencia e investigación Agraria*, volume (40), número (2), páginas (419-428).

[18] GENESIS ENERGÉTICA S.A. "Pequena Central Hidrelétrica São Francisco - Especificações Técnicas". Disponível em: <<http://www.silea.com.br/genesis/a-hidreletrica>>. Acesso em: 20 set. 2018.