

AVALIAÇÃO DA SEPARABILIDADE DE FITOFISIONOMIAS DO BIOMA CERRADO NO PARNA BRASÍLIA MEDIANTE TÉCNICAS DE CLASSIFICAÇÃO DE IMAGENS SAR SENTINEL-1 (BANDA C)

Andres Barrera Vilarmau¹, Jales de Freitas Bussinguer¹

¹PPGGAG, Universidade de Brasília, Brasília/DF, andres.vilarmau@aluno.unb.br, jales.bussinguer@aluno.unb.br

RESUMO

O estudo aborda a classificação temática de imagens SAR para a discriminação de três formações do Cerrado: Florestal, Savânica e Campestre. Foram analisadas as distribuições de intensidade e matriz de covariância C2 de produtos SLC Sentinel-1. A análise exploratória da média e do desvio padrão da série temporal de imagens indicou sobreposições importantes das classes em três recortes temporais distintos. As classificações supervisionada e não-supervisionada, utilizando dados do período seco, apresentou a melhor separabilidade de classes observada. Tais classificações retornaram exatidões acima de 71.3% para os algoritmos RF e K-Means. Para SVM, a exatidão máxima obtida foi de 52.39%. A principal hipótese levantada é que a performance dos classificadores é afetada indiretamente pelo efeito de saturação do sinal da banda C para as áreas estudadas.

Palavras-chave — matriz de covariância, Algoritmos de classificação, Cerrado, Banda C, Sentinel-1.

ABSTRACT

The study addresses the thematic classification of SAR images for the discrimination of three Cerrado formations: Florestal, Savânica and Campestre. The intensity distributions and C2 covariance matrix of SLC Sentinel-1 products were analyzed. The exploratory analysis of the mean and standard deviation of the time series of images indicated important overlaps of the classes in three different time frames. The supervised and unsupervised classifications, using data from the dry period, presented the best class separability observed. Such classifications returned accuracies above 71.3% for the RF and K-Means algorithms. For SVM, the maximum accuracy obtained was 52.39%. The main hypothesis raised is that the performance of the classifiers is indirectly affected by the saturation effect of the C-band signal for the studied areas.

Key words — covariance matrix, Classification Algorithms, Cerrado, Band C, Sentinel-1.

1. INTRODUÇÃO

O Bioma Cerrado ocorre principalmente no planalto central, ocupando aproximadamente 23% do território brasileiro (2.036.448 km² de extensão) [1]. Devido à riqueza e ao elevado grau de destruição atual, o bioma foi incluído na lista dos 34 hotspots da biodiversidade mundial [2].

O Sensoriamento Remoto permite o monitoramento de mudanças temporais no Bioma. A classificação permite extrair informações temáticas das imagens orbitais, a partir da aplicação de algoritmos matemáticos que categorizam a matriz de dados. Diante das limitações dos sensores passivos em áreas com cobertura recorrente de nuvens, como acontece no cerrado na época de chuvas (de outubro até março), os sensores SAR (Synthetic Aperture Radar) surgem como uma alternativa viável para o monitoramento contínuo de biomas, pois têm baixa interferência da atmosfera e fornecem dados sobre a geometria, densidade e as características dielétricas dos alvos na superfície. Contudo, poucos produtos têm sido gerados em escala global com dados SAR, destacando-se o mapa mundial de floresta/não-floresta produzido a partir dos dados ALOS-PALSAR (Banda L) [3]. Nesse sentido, Hansen et al. (2020) avaliaram o potencial dos dados do Sentinel-1 para a separação entre florestas e não-florestas por meio de algoritmos de classificação supervisionada em diferentes regiões do mundo, obtendo acurácias superiores a 87% [4].

Neste contexto, o objetivo do presente trabalho foi avaliar o potencial das imagens Sentinel-1, analisar intervalos temporais climáticos e utilizar três métodos de classificação sobre dados polarimétricos para discriminar as grandes formações vegetais do Cerrado localizadas no PARNA de Brasília.

2. MATERIAL E MÉTODO

2.1. Área de estudo

A área de estudo compreende o Parque Nacional de Brasília (PNB), localizado no Distrito Federal, dentro do Bioma Cerrado (**Figura 1**). Com área aproximada de 42.355 ha abrange as principais fitofisionomias deste Bioma [5].

O produto ESA WorldCover, proporciona uma classificação temática para o PNB com uma resolução de 10 metros para o ano de 2020 baseada em dados Sentinel-1 e Sentinel-2 [6]. As formações vegetais que estão presentes na área de estudo são: Florestal (FF), Savânica (FS) e Campestre (FC), identificadas pela WorldCover como *Tree cover*, *Shrubland* e *Grassland*, respectivamente. Estas formações vegetais do Cerrado são definidas como: formação campestre, no qual se encaixam áreas com predomínio do estrato herbáceo e algumas espécies arbustivas; formação savânica, com presença em diferentes proporções de arbustos e árvores, sem chegar a formar um dossel contínuo, sobre um estrato graminoso; e florestal, com predomínio de espécies arbóreas que formam um dossel contínuo ou descontínuo ao redor de corpos hídricos [7].

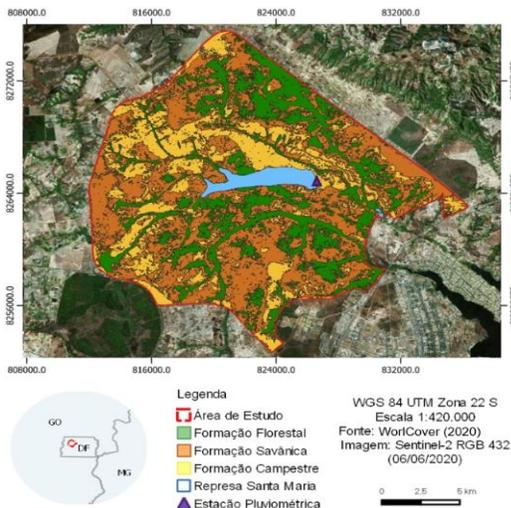


Figura 1. Área de estudo e feições de vegetação.

2.2. Abordagem de análise

O fluxo metodológico, iniciou-se pela caracterização dos períodos úmidos e secos no PNB, para o período de análise (setembro de 2019 a setembro de 2020). A partir dos dados pluviométricos da estação ETE-Paranoá, localizada dentro do PNB, foi definido como período úmido de 10/2019 até 05/2020, e como período seco o mês de 09/2019 e o período de 06/2020 até 09/2020. A partir dessa definição, fez-se a aquisição de imagens SAR Sentinel-1. Para o período entre 11/09/2019 e 29/09/2020 foram adquiridas 33 cenas. As imagens correspondem aos produtos *Single Look Complex* (SLC) da constelação Sentinel-1, que opera na banda-C com frequência de 5.4GHz (~5.6cm) em órbita descendente, adquirindo imagens pelo método *Interferometric Wide Swath* (IW) nas polaridades VV e VH.

As cenas foram submetidas a uma rotina de pré-processamento em lote no *software* SNAP 8.0. A rotina implementa atualização do arquivo de órbita, calibração radiométrica com preservação dos dados complexos; processamento de *deburst*; filtragem polarimétrica de ruído *speckle* (Refined Lee 5x5); correção geométrica e recorte das cenas. O produto final desse fluxo de processamentos é a matriz de covariância C2 para cada imagem da série temporal. As componentes da diagonal principal da matriz C2, C11 e C22 exprimem a intensidade do sinal nos canais VV e VH respectivamente e foram admitidas como parâmetros de análise para este estudo.

Em seguida, foram processados três conjuntos de rasters que representam a média e o desvio padrão temporal dos parâmetros C11 e C22 da série de imagens. O primeiro conjunto engloba toda a série temporal. O segundo conjunto refere-se apenas ao período úmido e o terceiro conjunto refere-se ao período seco na região de estudo. Tais estatísticas e recortes temporais foram adotados como atributos de análise deste estudo, partindo da hipótese de que estes apresentam potencial de separabilidade entre as classes estudadas.

A partir dos raster produzidos anteriormente, foi realizada uma análise exploratória dos dados por meio de gráficos de

distribuição da média e o desvio padrão para os alvos estudados (FF, FS e FC), para os três períodos temporais analisados e para cada uma das variáveis polarimétricas avaliados (C11 e C22).

Após a seleção dos melhores atributos, foram utilizados para avaliar a separabilidade de fitofisionomias dois métodos supervisionados: Support Vector Machines (SVM) e Random Forest Classifier (RFC) e um não supervisionado: K-Means (KM). As classificações foram realizadas, primeiro, considerando apenas o parâmetro C11 e em segundo, os parâmetros C11 e C22. Os produtos resultantes das classificações foram comparados ao mapa de vegetação do ESA WorldCover (2020), admitido como verdade de campo para este estudo.

Para a análise do desempenho dos classificadores foram utilizados o índice de exatidão global e o índice Kappa, tanto para o conjunto de classes como para cada classe individualmente. Tais dados foram gerados no *software* QGIS 3.16.10 com o *plugin* SCP. A exatidão global foi calculada a partir da matriz de confusão e expressa em termos de porcentagem em que, quanto mais próximo do 100%, melhor é o resultado da classificação. A qualificação dos índices Kappa foi realizada segundo a escala proposta por Landis e Koch [8].

3. RESULTADOS

A **Figura 2** apresenta a distribuição média e o desvio padrão dos parâmetros C11 e C22, para os três períodos de tempo estudados (série completa, período úmido e período seco), para os quatro alvos analisados presentes na área de estudo. Para todos os intervalos temporais estudados, os parâmetros C11 e C22 são maiores para a formação florestal, intermediários para a formação savânica e os menores valores são observados para a formação campestre. Para o alvo “água”, todos os parâmetros analisados e em todos os intervalos temporais objeto de estudo apresentaram os menores valores de retroespalhamento. A distribuição da água não apresenta sobreposição com o resto de alvos.

Na **Tabela 1** são apresentadas numericamente as correlações entre as distribuições da média e do desvio padrão para os distintos alvos estudados, para os três intervalos temporais e as duas variáveis analisadas: C11 e C22. Valores menores de correlação permitem melhor separação entre os alvos, o que melhora a classificação. Para facilitar a análise, sobre cada fila da **Tabela 1** se aplicou uma classificação de cores, sendo: verde a menor correlação (maior separabilidade) e vermelha a maior correlação (menor separabilidade) para os dois alvos comparados em cada fila.

É importante salientar que, devido ao comportamento espectral da água ser diferenciado do resto dos alvos estudados e que este apresenta uma separabilidade evidente em todos os parâmetros conforme indicado na **Figura 2**, serão comparadas unicamente as formações vegetais entre si.

A partir da análise dos resultados da **Figura 2** e da **Tabela 1**, foi observado que a distribuição obtida a partir do desvio padrão, para todos os parâmetros analisados, apresenta correlação entre os alvos (valores acima de 0.22), ou seja, existe sobreposição das distribuições. Dessa forma, este parâmetro

não permite uma boa separação das formações vegetais estudadas. Além disso, para os três períodos temporais avaliados, o desvio padrão apresentou uma maior correlação das distribuições que a média, para todos os alvos comparados (Tabela 1). Por estes motivos foi descartado o uso deste parâmetro na classificação.

| C11 | Completo | | Est. Seca | | Est. Chuvosa | |
|---------|----------|------|-----------|------|--------------|------|
| | média | std | média | Std | média | std |
| FF-FC | 0,41 | 0,59 | 0,32 | 0,68 | 0,46 | 0,91 |
| FC-FF | 0,07 | 0,22 | 0,02 | 0,28 | 0,13 | 0,75 |
| FC-Água | 0,01 | 0,00 | 0,02 | 0,07 | 0,02 | 0,03 |
| FS-FF | 0,54 | 0,68 | 0,42 | 0,76 | 0,64 | 0,93 |
| FS-Água | 0,01 | 0,01 | 0,02 | 0,00 | 0,02 | 0,01 |
| FF-Água | 0,02 | 0,01 | 0,03 | 0,01 | 0,02 | 0,01 |

| C22 | Completo | | Est. Seca | | Est. Chuvosa | |
|---------|----------|------|-----------|------|--------------|------|
| | média | std | média | Std | média | std |
| FF-FC | 0,65 | 0,79 | 0,33 | 0,76 | 0,70 | 0,99 |
| FC-FF | 0,10 | 0,44 | 0,08 | 0,34 | 0,29 | 0,98 |
| FC-Água | 0,04 | 0,04 | 0,03 | 0,02 | 0,05 | 0,02 |
| FS-FF | 0,57 | 0,75 | 0,40 | 0,76 | 0,71 | 0,99 |
| FS-Água | 0,03 | 0,03 | 0,03 | 0,03 | 0,04 | 0,02 |
| FF-Água | 0,04 | 0,04 | 0,03 | 0,03 | 0,05 | 0,02 |

Tabela 1. Correlação das distribuições de Formação campestre (FC), Formação savânica (FS, Formação florestal (FC) e água para as variáveis C11 e C22.

De forma geral, para as distribuições da média (Figura 2), a menor sobreposição é obtida entre as FF e FC. A maior sobreposição (> correlação) é encontrada entre as FF e FS.

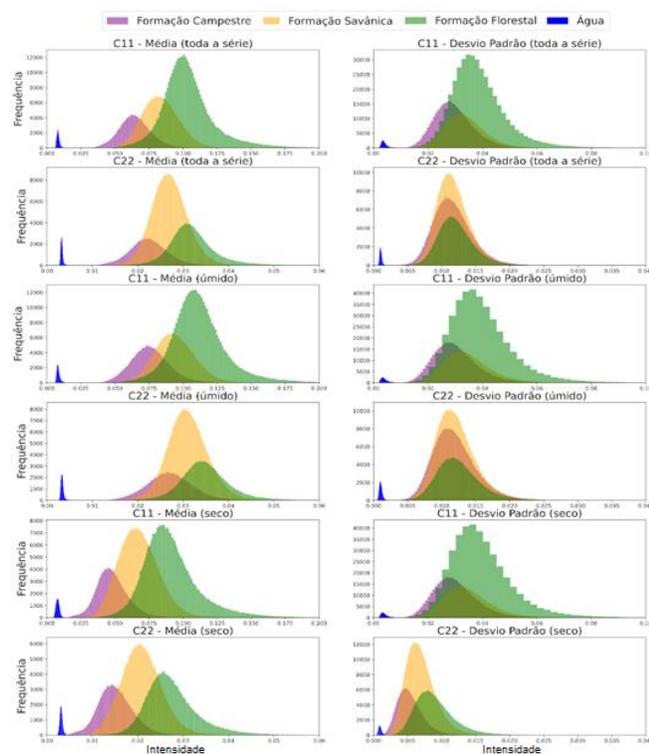


Figura 2. Distribuição da média e desvio padrão para os alvos estudados de C11 e C22.

Já as FC e FS apresentam uma sobreposição menor que as FS-FF. Esta diferença na sobreposição entre FC-FS e FS-FF

resulta em maior separabilidade das classes para a variável C11 que para C22, apresentando uma diferença média entre correlação FC-FS e FS-FF para os três períodos temporais avaliados de 0.13 para C11 e de 0.03 para C22.

A correlação é maior no conjunto de dados da estação chuvosa, para ambos os parâmetros C11 e C22, de modo que a separabilidade entre os alvos de estudo é comprometida. Nesse caso, o uso dos atributos desse período foi descartado.

A melhor separabilidade entre todas as fitofisionomias foi observada no período seco, para ambos os parâmetros C11 e C22. Dessa forma, o atributo escolhido para a classificação foi a média temporal do período seco.

O resultado do índice kappa para cada classe individualmente (Tabela 2) apresentou uma concordância perfeita na classificação da FF (0,85) com o classificador SVM C11 e para água (0,81) com a classificação RF C11-C22. Já as formações FC e FS alcançaram um nível de concordância Moderado (entre 0,56 e 0,61) com as classificações RF C11-C22, KM C11 e KM C11-C22.

| Classificador | FC | FS | FF | Água |
|---------------|-------|-------|------|------|
| SVM C11 | 0,22 | 0,19 | 0,85 | 0,44 |
| SVM C11 C22 | -0,02 | -0,38 | 0,68 | 0,21 |
| RF C11 | 0,51 | 0,37 | 0,47 | 0,63 |
| RF C11 C22 | 0,61 | 0,56 | 0,55 | 0,81 |
| KM C11 | 0,57 | 0,59 | 0,65 | 0,66 |
| KM C11 C22 | 0,60 | 0,61 | 0,65 | 0,69 |

Tabela 2. Resultados do índice Kappa, por classe estudada, para cada classificação avaliada.

O índice Kappa global das classificações (Tabela 3) apresentou índices próximos para as classificações RF C11-C22 (0,58), KM C11 (0,61) e KM C11-C22 (0,62). Tanto RF C11 (0,46) como RF C11-C22 (0,58) alcançaram um nível de concordância Moderado. O pior desempenho foi com a classificação SVM C11 (0,32, concordância Razoável), e SVM C11-C22, que alcançou uma concordância Mínima.

| Classificador | Classificação Kappa | Exatidão Total (%) |
|---------------|---------------------|--------------------|
| SVM C11 | 0,32 (razoável) | 52,39 |
| SVM C11 C22 | 0,03 (mínima) | 27,31 |
| RF C11 | 0,46 (moderada) | 63,10 |
| RF C11 C22 | 0,58 (moderada) | 71,30 |
| KM C11 | 0,61 (substancial) | 73,05 |
| KM C11 C22 | 0,62 (substancial) | 74,28 |

Tabela 3. Resultados do índice Kappa e Exatidão total (PT) de cada classificação.

Os classificadores K-Means (C11 e C11-C22) e Random Forest (C11-C22) foram os que apresentaram o melhor desempenho (Tabela 3), porém somente a classificação com K-Means alcançou uma concordância Substancial ($\geq 0,61$) segundo o índice kappa.

4. DISCUSSÃO

A maior sobreposição acontece na época de chuvas para o parâmetro C22 entre as formações savânica e florestal. Ambas formações podem alcançar valores de biomassa acima de 30 Mg.ha⁻¹ [9], sendo que a saturação do sinal, para a banda

C, acontece com valores de biomassa entre 30-50 Mg.ha⁻¹ [10].

A dificuldade na separação das formações savânica e campestre, na época chuvosa, está na predominância do estrato herbáceo em muitas áreas de ambas as formações [11]. Isto confere às classes uma estrutura semelhante e em razão disso, um padrão de retroespalhamento do sinal SAR similar.

Considerando apenas o período seco na área de estudo, a sobreposição da distribuição das classes é significativa. A principal hipótese relacionada a essa sobreposição é a limitação de penetração da banda C no dossel vegetal das fitofisionomias estudadas.

Cabe destacar a limitação das estatísticas adotadas para representar o comportamento temporal dos alvos. O desvio padrão se mostrou uma estatística frágil em separar os alvos, indicando que os desvios em torno da média seguem um mesmo padrão de distribuição. Em relação à média, observou-se que há uma proximidade entre os valores dos alvos estudados, o que pode dificultar a classificação temática utilizando apenas esta estatística, retornando baixas acurácias.

Quanto ao desempenho dos classificadores, um fator determinante para o sucesso da classificação é a quantidade de atributos utilizados. O aumento do número de parâmetros, melhorou o desempenho do Random Forest e do K-Means. Contudo, ao avaliar a classificação dos índices kappa das respectivas classificações, notou-se que estas ocupam uma região limiar entre a exatidão moderada e substancial. Tal cenário indica uma incerteza na categorização da qualidade dos resultados.

Hansen et al. [4] estudaram o comportamento de duas classes: floresta e não-floresta. A classe não-floresta engloba alvos como agricultura, solo exposto, ambientes urbanos, água e vegetações não-florestais. Nesse caso, a separabilidade das classes era nítida, com baixa sobreposição entre as classes, levando a acurácias superiores a 87% em diferentes biomas ao redor do mundo. No caso específico deste trabalho, a tentativa é diferenciar três alvos vegetais com características semelhantes em dados da banda C, o que dificulta a separabilidade dos alvos estudados. Nesse caso, cabe destacar que, para o SVM, o uso de dois parâmetros (C11-C22) piorou o comportamento do classificador, acarretando uma maior ocorrência de regiões classificadas como “água”, porém elucidar os motivos desse comportamento fica fora do escopo deste trabalho.

5. CONCLUSÕES

Os resultados deste estudo mostram que os dados de intensidade do sinal derivados de produtos SLC da constelação Sentinel-1 apresentam limitações na discriminação das fitofisionomias de Cerrado encontradas no PNB. A principal hipótese é de que a saturação do sinal da banda C para estes alvos limita a separabilidade das classes.

O período temporal mostrou-se um fator determinante para o desempenho das classificações, uma vez que o maior potencial de separabilidade das classes foi observado em um

período específico da série temporal. Além disso, a influência do regime de chuvas pode potencializar o efeito de saturação do sinal da banda C para as fitofisionomias estudadas.

Os resultados mostram que os algoritmos de classificação, K-Means e Random Forest principalmente, quando aplicados nas imagens Sentinel-1, podem ser utilizados como alternativas ao uso de imagens ópticas, caso não estejam disponíveis ou em períodos de alta cobertura de nuvens.

Para trabalhos futuros é recomendável a exploração da classificação de imagens Sentinel-1 por meio de técnicas de segmentação de imagens em uma abordagem de classificação orientada a objetos (GEOBIA).

8. REFERÊNCIAS

- [1] IBGE. Mapa de Biomas do Brasil - Primeira Aproximação. 2004.
- [2] Mittermeier, R. A., Gil, P. R., Hoffmann, M., Pilgrim, J., Brooks, T., Mittermeier, C. G., Lamoureaux, J. & Fonseca, G. A. G. Hotspots revisited. Earth's biologically richest and most endangered terrestrial ecoregions.. CI, Washington. Pp. 15-96. 2005.
- [3] Shimada, M.; Itoh, T.; Motooka, T.; Watanabe, M.; Shiraishi, T.; Thapa, R.; Lucas, R. Nuevos mapas globales de bosques y no bosques a partir de datos de ALOS PALSAR (2007–2010). Ambiente Sens. Remoto. 155, 13–31. 2014.
- [4] Hansen, J.N.; Mitchard, E.T.A.; King, S. Assessing Forest/Non-Forest Separability Using Sentinel-1 C-Band Synthetic Aperture Radar. Remote Sensing. 12, 1899. <https://doi.org/10.3390/rs12111899>. 2020.
- [5] Plano de Manejo do PARNA Brasília. Convenio IBAMA-FUNATURA. 1996.
- [6] WorldCover. Worldwide land cover mapping. 2020. Disponível em: <https://esa-worldcover.org/en>. Acesso em: novembro 2021.
- [7] Ribeiro, J.F.; Walter, B.M.T.; Fitofisionomias do bioma Cerrado. In: Sano, S.M.; Almeida, S.P. (Ed.). Cerrado: ambiente e flora. Planaltina, DF: Embrapa-CPAC, cap 3, p.87-166. 1998.
- [8] Landis, J.; Koch, G. The measurement of observer agreement for categorical data. Washington, USA. Biometrics, v. 33, n. 1, p. 159-174. 1977
- [9] Roquette, J. Distribuição da biomassa no Cerrado e a sua importância na armazenagem do carbono. Ciência Florestal. 28. 1350. 10.5902/1980509833354. 2018.
- [10] Quegan, S.; Le Toan, T.; Yu, J. J.; Ribbes F.; Floury, N. Multitemporal ERS SAR analysis applied to forest mapping. *IEEE T&RS*, vol. 38, no. 2, pp. 741-753, doi: 10.1109/36.842003. 2000.
- [11] Plano de Manejo do Parque Nacional de Brasília. Ed. ICMBio. Cap. 5.2.1.1. Vegetação, pág 5.54-5.60. 1998