

Classificação Automática de Cobertura de Solo com Séries Temporais do Sentinel-2 e do MapBiomas

João Leonardi da Silva Melo, Dimmy Karson Soares Magalhães,
Javier Ernesto Kolodziej e Eduardo Vínicius Kuhn

Resumo— Este trabalho apresenta um método de classificação automática de cobertura de solo a partir de séries temporais obtidas pelo Sentinel-2 L2A e dados de referência extraídos do MapBiomas (Coleção 10). Um classificador *Random Forest* é empregado com validação cruzada espacial por blocos em três propriedades da mesorregião oeste do Paraná, com rotações de culturas distintas. Os resultados (acurácia global OA 0,7625 e coeficiente Kappa κ 0,6195) quantificam a concordância com o MapBiomas sob validação espacial por blocos (não a acurácia de campo). O confronto com os calendários dos proprietários, usado como checagem de coerência temporal, indica retenção de soja superior a 84% nas três áreas à medida que o empilhamento temporal se consolida, sugerindo que o classificador captura a assinatura espectro-temporal da rotação soja–cereal de inverno.

Palavras-Chave— Classificação de cobertura do solo, MapBiomas, Sensoriamento remoto, Sentinel-2, Séries temporais.

Abstract— This work presents a land cover classification method using Sentinel-2 L2A time series and MapBiomas (Collection 10) reference data. A Random Forest classifier is employed with spatial block cross-validation on three properties in the western region of Paraná state, Brazil, with distinct crop rotations. Results (overall accuracy OA 0.7625 and Kappa coefficient κ 0.6195) quantify agreement with MapBiomas under spatial block cross-validation (not pixel-level field accuracy). Landowner crop calendars, used as a temporal coherence check, showed soybean retention above 84% across the three areas as the temporal stacking consolidates, suggesting that the classifier captures the integrated temporal signature of the soybean–winter cereal rotation.

Keywords— Land cover classification, MapBiomas, Remote sensing, Sentinel-2, Time series.

I. INTRODUÇÃO

O Brasil é um dos maiores produtores e exportadores de commodities agrícolas, com produção estimada de 176 milhões (soja) e 140 milhões (milho) de toneladas na safra 2025/26 [1]. A classificação de cobertura de solo é instrumento essencial de gestão territorial, sobretudo onde ciclos fenológicos diversos tornam o monitoramento convencional ineficiente [2]. O sensoriamento remoto por satélite permite

observação sistemática em larga escala; séries temporais caracterizam padrões espectro-temporais entre classes ao longo do ciclo anual [3]. O Sentinel-2 (ESA/Copernicus) oferece imagens multiespectrais com resolução de 10 m e revisita de 5 dias [4], adequadas ao monitoramento agrícola em escala de parcela.

Neste trabalho, apresenta-se um método de classificação de cobertura de solo, utilizando descritores espectro-temporais derivados de imagens do Sentinel-2 e de referência do MapBiomas Coleção 10. Especificamente, objetiva-se: (i) construir e avaliar um vetor de *features* espectro-temporais para classificação via *Random Forest* com validação espacial por blocos; (ii) analisar o desempenho por classe em três propriedades da mesorregião oeste do Paraná com rotações de culturas distintas; (iii) confrontar as predições com calendários agrícolas documentados pelos proprietários, avaliando a coerência temporal da classificação. O restante deste trabalho está organizado como segue. A Seção II descreve os materiais e métodos. A Seção III apresenta e discute os resultados obtidos. E, a Seção IV sintetiza as considerações finais.

II. MATERIAIS E MÉTODOS

A. Áreas de estudo

Três propriedades privadas na mesorregião oeste do Paraná (Fig. 1) são consideradas. A Propriedade A, localizada em (25°05'S, 53°32'W), possui aproximadamente 30 ha e apresenta cinco classes de cobertura segundo o MapBiomas (Coleção 10): Soja (55,0%, $n = 1679$), Formação Florestal (14,2%), Mosaico de Agricultura e Pastagem (14,2%), Outras Lavouras Temporárias (13,2%) e Área não Vegetada (3,4%), totalizando $n = 3010$ amostras. As Propriedades B (25°44'S, 53°31'W) e C (25°08'S, 53°31'W), com aproximadamente 40 ha e 71 ha, respectivamente, correspondem a áreas monoclasse de Soja segundo o MapBiomas, com todos os pixels rotulados

TABELA I
HISTÓRICO DE CULTURAS DOCUMENTADO PELOS PROPRIETÁRIOS DAS TRÊS ÁREAS DE ESTUDO (2024–2026).

Período	Propriedade		
	A	B	C
Mai.–Set./2024	Trigo	Trigo	Milho
Out./2024–Fev./2025	Soja	Milho+Soja	Soja
Mar.–Abr./2025	Pousio	Nabo/T. Mor.	Milho
Mai.–Set./2025	Aveia	Trigo	Milho
Out./2025–Jan./2026	Soja	Milho+Soja	Soja

João Leonardi da Silva Melo e Dimmy Karson Soares Magalhães estão vinculados ao ICEV Instituto de Ensino Superior, Teresina, PI, Brasil (e-mail: joao_leonardi.melo@somosicev.com; dimmy.magalhaes@somosicev.com).

Javier Ernesto Kolodziej é docente e pesquisador visitante no programa de Pós-graduação em Tecnologias em Biociências da Universidade Tecnológica Federal do Paraná (UTFPR), Toledo, PR, Brasil e está vinculado ao Instituto Misionero de Materiales, UNaM-CONICET, Oberá, Argentina (e-mail: javierkolodziej@utfpr.edu.br).

Eduardo Vínicius Kuhn está vinculado ao Programa de Pós-graduação em Tecnologias em Biociências (PPGBio) da Universidade Tecnológica Federal do Paraná (UTFPR), Toledo, PR, Brasil, e ao Programa de Pós-graduação em Ciência da Computação (PPGComp) da Universidade Estadual do Oeste do Paraná (UNIOESTE), Cascavel, PR, Brasil (e-mail: kuhn@utfpr.edu.br).

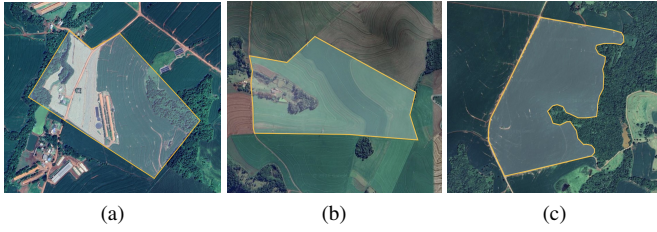


Fig. 1. Áreas de estudo consideradas na mesorregião oeste do Paraná. (a) Propriedade A: (25°05'S, 53°32'W). (b) Propriedade B: (25°44'S, 53°31'W). (c) Propriedade C: (25°08'S, 53°31'W). Fonte: Google Earth ©.

como tal, totalizando $n = 4011$ e $n = 7105$ amostras, respectivamente.

B. Conjunto de dados

Features espectro-temporais são derivadas de 22 imagens multiespectrais obtidas pelo Sentinel-2 L2A (nível de refletância de superfície), adquiridas entre maio de 2024 e janeiro de 2026, nas bandas B02 (490 nm), B03 (560 nm), B04 (665 nm) e B08 (842 nm), com resolução de 10 m. Os rótulos de referência provêm do MapBiomias Coleção 10, ano-base 2023, com resolução original de 30 m e reamostrados para 10 m por interpolação *Nearest Neighbor*. Adotou-se o vizinho mais próximo para preservar a natureza categórica dos rótulos. (Métodos de interpolação contínua gerariam classes intermediárias inválidas, embora aumentem o risco de pixels mistos na escala de 10 m.) Tais rótulos cumprem dois papéis no presente trabalho: alimentar o treinamento do classificador e servir como referência na validação cruzada interna (Seção II-D). Como ambos os usos compartilham a mesma fonte, adota-se também uma fonte independente de validação, as rotações de culturas documentadas pelos proprietários no período 2024–2026 (Tabela I), que permite confrontar as previsões com a cultura efetivamente presente em campo.

C. Engenharia de Features

O vetor de *features* é construído em três etapas. Primeiro, para cada pixel, os índices de vegetação NDVI [5], EVI [6] e SAVI [7] são calculados a partir das 22 datas de aquisição utilizando as bandas B02, B04 e B08 do Sentinel-2. Em seguida, extraem-se 14 estatísticas descritivas de cada série temporal (média, desvio padrão, mediana, mínimo, máximo, amplitude, IQR (intervalo interquartil), coeficiente de variação, assimetria, curtose e percentis 10, 25, 75 e 90), totalizando 42 *features*. Na segunda etapa, calculam-se média, desvio padrão, mínimo e máximo para cada banda espectral (B02, B03, B04 e B08), resultando em 16 *features*. Por fim, adicionam-se os 22 valores brutos de NDVI (um por data de aquisição), preservando o perfil temporal. Assim, cada pixel é representado por um vetor de 80 *features*.

D. Classificação e validação

O *Random Forest* [8], frequentemente empregado em sensoriamento remoto por sua tolerância a *features* correlacionadas

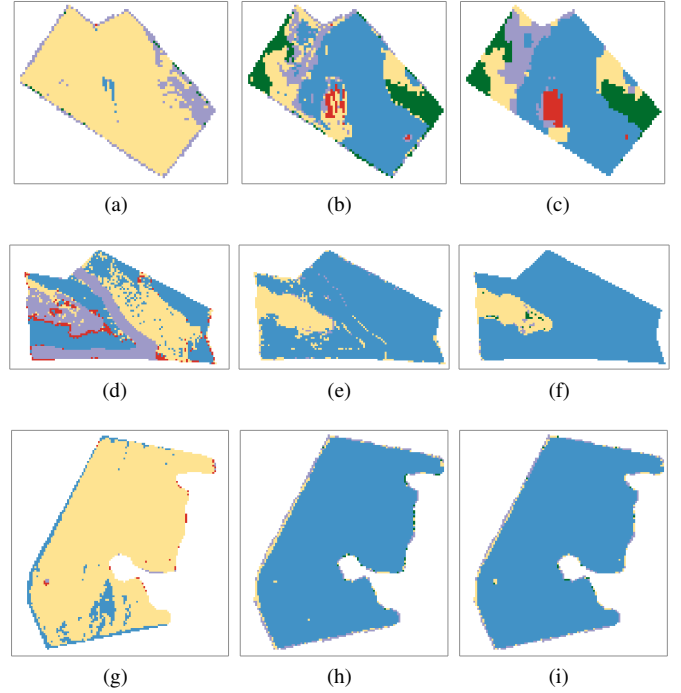


Fig. 2. Classificação de cobertura de solo de cada propriedade, a partir de imagens do satélite Sentinel-2 L2A (ESA/Copernicus), em três datas representativas (da esquerda para a direita), considerando formação florestal (verde), mosaico de agricultura-pastagem (amarelo), área não-vegetada (vermelho), soja (azul) e outras lavouras temporárias (roxo). (a)-(c) Propriedade A, avaliada em Jul./24, Abr./25 e Dez./25. (d)-(f) Propriedade B, avaliada em Jul./24, Mar./25 e Jan./26. (g)-(i) Propriedade C, avaliada em Jul./24, Abr./25 e Jan./26.

e baixo custo de ajuste [9], é configurado com 300 árvores, $\text{max_features}=\sqrt{p}$ e $\text{class_weight}=\text{balanced}$, e treinado com os rótulos do MapBiomias (Seção II-B). Adotaram-se 300 árvores e $\text{max_features}=\sqrt{p}$ ($p = \text{n}^\circ$ de *features*) conforme a prática usual em RF [9]; $\text{class_weight}=\text{balanced}$ compensa o desbalanceamento na Propriedade A. A busca em grade com validação espacial por blocos não produziu ganhos fora da amostra. Para considerar a autocorrelação espacial [10], emprega-se *StratifiedGroupKFold* com 5 *folds* cujos grupos são blocos de 10×10 pixels (100 m). As previsões dos *folds* são agregadas via cross_val_predict sobre o conjunto completo ($n = 3010$ amostras), do qual derivam-se acurácia global (OA), coeficiente Kappa (κ), F1 (macro e ponderada) e IoU (*Intersection over Union*) por classe; como há agregação, e não média, não se reporta variância entre partições. As métricas na Propriedade A ($n = 3010$) medem a concordância com os rótulos do MapBiomias (não a acurácia de campo). Os calendários fornecem apenas uma verificação temporal complementar.

E. Validação com o histórico dos proprietários

A validação independente segue um protocolo de empilhamento temporal cumulativo: o classificador, treinado uma

única vez com o vetor de features completo (22 datas), é aplicado a cada propriedade em T etapas, sendo o vetor reconstruído na t -ésima etapa apenas com as imagens entre t_1 e t . Em cada etapa, calcula-se a retenção de soja, definida como a fração dos pixels rotulados como soja pelo MapBiomias que o classificador mantém nessa classe. Essa métrica avalia como a evolução temporal das features afeta a coerência entre predição e referência ao longo do ciclo agrícola. Para as Propriedades B e C, monoclasse de soja, a retenção equivale à proporção total de pixels classificados como soja; para a Propriedade A, restringe-se ao subconjunto de pixels rotulados como soja ($n = 1679$). Para visualização, três datas por propriedade são selecionadas (Tabela I), utilizando-se a aquisição do Sentinel-2 mais próxima do ponto médio temporal de cada período. A coerência entre as predições e os calendários é analisada na Seção III. Esse protocolo não substitui uma avaliação pixel a pixel com verdade de campo independente.

TABELA II

MÉTRICAS DE CLASSIFICAÇÃO POR CLASSE E GLOBAIS (5 folds COM PARTIÇÃO ESPACIAL POR BLOCOS, $n = 3010$).

Classe	IoU	n amostras
Soja	0,8103	1679
Formação Florestal	0,6667	420
Área não Vegetada	0,3889	96
Outras Lavouras Temporárias	0,3599	387
Mosaico Agric.–Pastagem	0,2568	428
Métricas globais		
Acurácia global (OA)	0,7625	
Coefficiente Kappa (κ)	0,6195	
F1 macro	0,6386	
F1 ponderado	0,7550	
mIoU (média dos IoUs)	0,4965	

III. RESULTADOS E DISCUSSÃO

A Tabela II apresenta as métricas globais (OA 0,7625, κ 0,6195, mIoU 0,4965) e os IoUs por classe, interpretados como concordância com o MapBiomias na Propriedade A (não como acurácia de campo). Os melhores desempenhos foram observados para Soja (IoU 0,81), associada à sua assinatura fenológica distinta no perfil NDVI e à predominância na área (55%), e Formação Florestal (0,67), consistente com sua estabilidade espectral ao longo do ciclo anual; os piores ocorreram para Mosaico de Agricultura e Pastagem (0,26) e Outras Lavouras Temporárias (0,36), possivelmente devido à heterogeneidade espectral intrínseca desses rótulos no MapBiomias.

A Fig. 2 apresenta os mapas em três datas representativas. Em Jul./2024, com cultivos de inverno em campo, a retenção de Soja foi baixa em todas as áreas (1%, 46% e 9% nas Propriedades A, B e C, respectivamente); a partir de Jan./2025, à medida que o empilhamento temporal incorporou os primeiros ciclos de Soja, a retenção convergiu para 84–98% e manteve-se em 91–98% nos ciclos subsequentes. Na Propriedade C, os períodos de milho foram classificados predominantemente

como Mosaico de Agricultura e Pastagem. O modelo distingue, portanto, a Soja dos demais cultivos e captura a assinatura temporal da rotação soja–cereal de inverno, embora agrupe as culturas de inverno (trigo, aveia, milho de inverno) sob classes não-Soja sem diferenciá-las, limitação esperada, já que essas culturas não constituem rótulos individuais no MapBiomias (Coleção 10) para a região.

IV. CONSIDERAÇÕES FINAIS

Apresentou-se um método de classificação de cobertura do solo com 80 *features* espectro-temporais derivadas de 22 imagens Sentinel-2, referência MapBiomias (Coleção 10) e validação espacial por blocos. Os resultados (OA 0,7625 e κ 0,6195) expressam a concordância com o MapBiomias sob validação espacial por blocos, com retenção de Soja superior a 84% nas três áreas após o empilhamento temporal, sugerindo que o método captura a assinatura espectro-temporal da rotação soja–cereal de inverno. As limitações incluem: (i) dependência circular entre treinamento e validação cruzada com o MapBiomias; (ii) defasagem temporal (rótulos 2023 vs. imagens 2024–2026); (iii) pixels mistos decorrentes da amostragem de 30 m para 10 m; e (iv) ausência de verdade de campo independente em nível de pixel. Trabalhos futuros podem incorporar rótulos de campo e arquiteturas temporais (LSTM/Transformers).

AGRADECIMENTOS

Os autores agradecem aos produtores Delacir Zanatta, Lucas Batistella e Carlos Luiz Daroda pela disponibilização dos calendários agrícolas com o histórico das culturas, bem como pelos dados necessários à identificação das áreas consideradas.

REFERÊNCIAS

- [1] COMPANHIA NACIONAL DE ABASTECIMENTO, “Acompanhamento da safra brasileira de grãos: Safra 2025/26: Sétimo levantamento,” CONAB, Brasília, Tech. Rep. no. 7, Apr. 2026, Acesso em: 22 maio 2026.
- [2] D. Arvor et al., “Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil,” *International Journal of Remote Sensing*, vol. 32, no. 22, pp. 7847–7871, Nov. 2011.
- [3] C. Pelletier et al., “Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas,” *Remote Sensing of Environment*, vol. 187, pp. 156–168, Dec. 2016.
- [4] M. Drusch et al., “Sentinel-2: ESA’s optical high-resolution mission for GMES operational services,” *Remote Sensing of Environment*, vol. 120, pp. 25–36, May 2012.
- [5] J. W. Rouse et al., “Monitoring vegetation systems in the great plains with ERTS,” in *Third Earth Resources Technology Satellite-1 Symposium*, ser. NASA SP-351, vol. 1, Washington, DC: NASA, 1974, pp. 309–317.
- [6] A. Huete et al., “Overview of the radiometric and biophysical performance of the MODIS vegetation indices,” *Remote Sensing of Environment*, vol. 83, no. 1–2, pp. 195–213, Nov. 2002.
- [7] A. R. Huete, “A soil-adjusted vegetation index (SAVI),” *Remote Sensing of Environment*, vol. 25, no. 3, pp. 295–309, Sep. 1988.
- [8] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [9] M. Belgiu and L. Drăguț, “Random forest in remote sensing: A review of applications and future directions,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, Apr. 2016.
- [10] D. R. Roberts et al., “Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure,” *Ecography*, vol. 40, no. 8, pp. 913–929, Aug. 2017.