

MAPEAMENTO DA CITRICULTURA IMAGENS SENTINEL – 2 E RANDOM FOREST: O EXEMPLO DE CASA BRANCA

Victória Hellena Matusевичius e de Castro ^{1,3}, Taya Cristo Parreiras ²,
Édson Luis Bolfe ^{2,3}, Luiz Eduardo Vicente ⁴

¹ Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, CEP 13083-875, Campinas – SP, v245192@dac.unicamp.br; ² Instituto de Geociências, Universidade Estadual de Campinas, CEP 13083-855, Campinas – SP, tayacristo1@gmail.com; ³ Embrapa Agricultura Digital, CEP 13083-886, Campinas – SP, edson.bolfe@embrapa.br; ⁴ Embrapa Meio Ambiente, CEP 13820-000, Jaguariúna – SP, luiz.vicente@embrapa.br

RESUMO

O Brasil se destaca mundialmente como o segundo maior produtor de citros e o maior produtor de suco de laranja. Apesar da relevância econômica, o mapeamento dessas culturas ainda é um desafio, sobretudo para o sensoriamento remoto voltado para o monitoramento de larga escala. Portanto, este estudo teve como objetivo mapear a produção de citricultura do município de Casa Branca, São Paulo, na safra 2021/2022. Para isso, foi proposta uma metodologia baseada em medianas mensais do conjunto multispectral de bandas do Sentinel-2, auxílio de dados de campo e algoritmo *Random Forest*. A metodologia proposta foi capaz de detectar as áreas ocupadas com citros com precisão de 88%, estimando a área de produção com uma diferença de cerca de 1 mil ha em relação aos dados de Produção Agrícola Municipal (PAM).

Palavras-chave - Citros, Sensoriamento Remoto, Random Forest.

ABSTRACT

Brazil stands out worldwide as the second-largest producer of citrus and the largest producer of orange juice. Despite the economic relevance, mapping these crops is still a challenge, especially for remote sensing aimed at large-scale monitoring. Therefore, this study aimed to map the citrus production in Casa Branca, São Paulo state, in the 2021/2022 crop season. Therefore, a methodology was proposed based on monthly medians of the Sentinel-2 multispectral set of bands, aided by field data and the Random Forest algorithm. The proposed methodology was able to detect the areas occupied with citrus with an accuracy of 88%, estimating the production area with a difference of about 1,000 ha compared to the reference data.

Key words - Citros, Remote Sensing, Random Forest.

1. INTRODUÇÃO

A citricultura, que engloba os cultivos de laranja, limão e tangerina, teve início no Brasil com os jesuítas por volta de 1530, com a produção de laranja nos estados da Bahia e São Paulo [1].

Contudo, o país despontou no cenário internacional apenas na década de 1980, quando as plantações de laranja na Flórida, atingidas pelas geadas, não conseguiam mais atender ao mercado consumidor e o Brasil começou a exportar, vindo a se tornar o maior produtor e exportador mundial. Devido às condições naturais favoráveis, boa capacidade logística e mão de obra barata, o estado de São Paulo solidificou-se, a partir da década de 1970, como principal estado produtor [2].

De acordo com os dados da pesquisa da Produção Agrícola Municipal (PAM), do Instituto Brasileiro de Geografia e Estatística (IBGE) [3], cerca de 6% da produção paulista de laranja é realizada no município de Casa Branca, que também produz 0,21% do limão e 8% da tangerina de todo o estado. Aproximadamente 85% da área do município é destinada à agricultura, sendo 5,40% dessa área destinada à citricultura, com rendimento médio da produção em 2021 de 55.000 kg/ha para laranja, 54.506 kg/ha para tangerina e 18.260 kg/ha para limão. As condições edáficas, climáticas e geomorfológicas de Casa Branca favoreceram o cultivo da citricultura. De acordo com dados disponibilizados pelo SEADE [4], o PIB agrícola corresponde a 17,7% do PIB municipal, movimentando 160 milhões de reais em 2019, sendo 44,3% desse valor proveniente do cultivo de citros.

O mapeamento da dinâmica espacial no cultivo de lavouras permanentes, em especial a citricultura, é importante para fornecer informações para a tomada de decisão no planejamento, gestão e manejo das áreas produtivas pelos produtores, gestores e consumidores [5-6]. Embora o sensoriamento remoto (SR) possua papel estratégico no acompanhamento dessa dinâmica, técnicas para mapeamento e monitoramento dessas lavouras ainda permanecem subavaliadas e preteridas em relação às lavouras anuais [7].

Neste estudo, objetivou-se analisar o potencial de uma série temporal anual de medianas mensais de bandas espectrais do sensor *Mustispectral Instrument* (MSI), dos satélites Sentinel-2 A e B, em combinação com o algoritmo *Random Forest* e dados de campo, para o mapeamento das áreas de produção citros do município de Casa Branca, SP.

2. MATERIAL E MÉTODOS

2.1. Área de estudo

Casa Branca é um município de 864 km², do estado de São Paulo, localizado no nordeste de São Paulo entre as coordenadas 47° 5' 17.258" W e 21° 48' 1.971" S (Figura 1). O solo mais abundante do município é o Latossolo Vermelho Amarelo, e 43,8% do município apresenta declividade inferior à 8% [8]. O clima é o subtropical úmido (Cwa) [9], a temperatura média anual é de 20,5°C [10], dentro de um intervalo ótimo de 23 a 32°C para o cultivo de citros [11]. De acordo com dados estimados do *Global Precipitation Measurements* (GPM), disponíveis via *Google Earth Engine* (GEE), o município apresentou média histórica de precipitação anual acumulada de 1.459 mm entre 2001 e 2020, dentro das exigências anuais da cultura, que variam entre 900 e 1.500 mm [11].

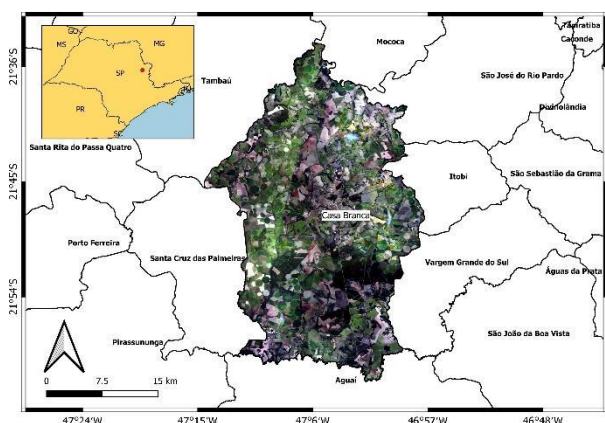


Figura 1. Localização do município de Casa Branca (SP). Imagem de fundo: Composição Landsat em cor verdadeira, referente à 07/2022.

2.2 Aquisição e processamento de dados

Com o objetivo de fomentar a discussão dos resultados, dados da sétima coleção do MapBiomas e da PAM foram obtidos para os anos de 1985, 1990, 1995, 2000, 2005, 2010, 2015, 2020 e 2021. Os mapas do MapBiomas foram adquiridos via GEE, e a área total ocupada por cada classe, em cada ano, foi extraída por meio da ferramenta *r-report*, no QGIS 3.10. Os dados referentes à área destinada à colheita de laranja, limão e tangerina foram obtidos da PAM [3].

As imagens Sentinel-2 MSI Level-2^a (classificação de cena e uma correção atmosférica aplicada a produtos de ortoimagem Nível-1C, em valores de reflectância do fundo da atmosfera) nas bandas 2, 3, 4, 5, 6, 7, 8, 8A, 9 11 e 12, também foram adquiridas via GEE, na forma de medianas mensais, entre setembro de 2021 e agosto de 2022. Em seguida, as 11 bandas de cada um dos 12 meses foram empilhadas para formar um arquivo final de 132 camadas. A estratégia de utilizar medianas mensais foi adotada para reduzir a dimensionalidade do conjunto de dados. Os vetores do Cadastro Ambiental Rural (CAR) [12] foram utilizados para concentrar a classificação em áreas rurais.

2.4 Classes mapeadas e conjunto de dados amostrais

Um total de 341 pontos amostrais foram coletados *in situ*, com suporte do aplicativo AgroTag [13], durante duas campanhas de campo, entre 28 e 29 de maio, e 17 e 18 de agosto, em 2022. De forma complementar, 179 novas amostras foram obtidas de forma remota, por meio de fotointerpretação, utilizando mosaicos mensais Planet (Imagem © 2022 Planet Labs Inc. Todos os usos estão sujeitos ao Contrato de Licença do Participante), de 4,7 m de resolução espacial, bem como considerando o perfil histórico do NDVI gerado no SatVeg [14].

Para obter um volume de amostras significativo e representativo, as classes e o número de pontos amostrais foram definidos considerando as mais recentes estimativas de área ocupada por cada tipo de uso, sendo os dados da PAM 2021 (lavouras anuais e permanentes) e Pesquisa da Extração Vegetal e da Silvicultura (PEVS, para silvicultura), ambos via portal Sidra [3, 15], e da 7ª coleção do MapBiomas (áreas de pastagens, mata nativa e corpos hídricos). A Tabela 1 mostra as classes definidas, as estimativas de área ocupada, bem como o número de pontos amostrais para cada classe.

Classe	Área (ha)	Amostras
Água	778	40
Vegetação Nativa	10.000	80
Lavouras Anuais	52.596	200
Pastagens	1.991	50
Silvicultura	6.226	50
Citros	14.658	100

Tabela 1. Estimativas de área ocupada e número de pontos amostrais por classe definida.

2.5 Classificação digital e análise de acurácia

O conjunto de dados foi aleatoriamente dividido em um subconjunto de treinamento (70%) e um de teste (30%), e o modelo de classificação foi gerado com algoritmo *Random Forest*, no R. No que diz respeito aos parâmetros do algoritmo, o número de variáveis por nó (mTry) foi definido como a raiz quadrada do número de variáveis [16], e o número de árvores (nTree) foi empiricamente definido como 300, observando a estabilização do erro a partir da adição de novas árvores. Para avaliar a performance do modelo, foi gerada a matriz de confusão, e foram consideradas as seguintes métricas: acurácia global (AG), índice de Kappa (K), além da acurácia balanceada por classe (PB), e a precisão do usuário (PU) e do produtor (PP). A área total estimada de citricultura no município foi também comparada com os dados da PAM 2021. Após a validação, um filtro de moda em janela 3 x 3 foi aplicado ao mapa final para minimizar o efeito *Salt and Pepper*.

3. RESULTADOS E DISCUSSÕES

3.1 Estimativas de área com produção de citricultura

A Figura 3 apresenta os dados referentes à área colhida de citros, levantados pela PAM para os anos de 1985, 1990, 1995, 2000, 2005, 2010, 2015, 2020 e 2021.

A irrigação, a adoção de melhores práticas de manejo e a maior densidade adotada no plantio culminaram com o aumento da produtividade e isso, associado ao aumento progressivo no custo da produção, resultaram na redução da área destinada à produção de citros em São Paulo entre 2001 e 2015, com migração principalmente para a produção de cana-de-açúcar [17]. Essa redução é observada em Casa Branca, onde a PAM indicou o decréscimo de cerca de 2.200 ha que deixaram de ser cultivados com citros entre 2000 e 2020. Embora pesquisas auto declaratórias como a PAM representem dados de referência, elas não são capazes de revelar a distribuição e a dinâmica espacial da produção agrícola, tarefa que cabe ao SR.

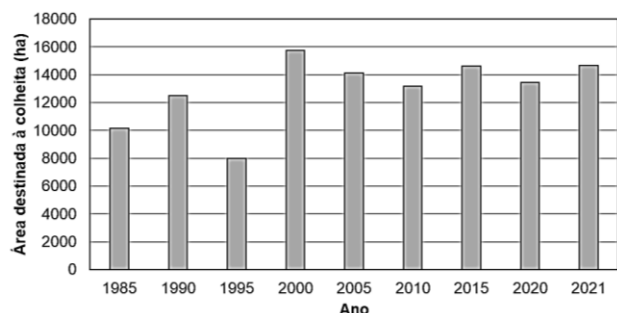


Figura 3. Área de produção de citros, em hectares, entre 1985 e 2021, levantada pela pesquisa de Produção Agrícola Municipal (PAM).

3.2 Mapeamento da citricultura e acurácia

A metodologia proposta estimou, com 90% de acurácia global, uma área de 13.559 ha de citricultura no município de Casa Branca, na safra 2021/2022, o que representa uma diferença absoluta de 1.099 ha em relação aos dados da PAM no mesmo ano. A precisão da classe e a precisão do produtor para a classe 88% e 76%, respectivamente (Tabela 2).

Classe	AG	VG	LA	PA	SI	CT	PU%
AG	10	0	0	0	0	0	100
VG	0	26	0	0	3	4	78
LA	0	0	52	2	2	2	89
PA	0	0	0	5	0	2	71
SI	0	0	0	0	18	0	100
CT	0	0	0	0	0	25	100
PP%	100	100	100	71	78	76	
PB%	100	97	97	85	89	88	
AG; K (%)			80; 97				

Tabela 2. Matriz de confusão e resultados obtidos pelo modelo de classificação com imagens Sentinel-2 e algoritmo Random Forest. Em que: AG (água); VG (vegetação nativa); LA (lavouras anuais); PA (pastagens); SI (silvicultura); CT (citros); PU (precisão do usuário); PP (precisão do produtor); PB (precisão balanceada).

A Figura 4 apresenta o resultado da classificação, por classe de uso e cobertura das terras, no município de Casa Branca (SP), bem como uma aproximação de um mosaico

mensal do Planet e uma fotografia tirada *in situ*, ambas de maio de 2022. Foram observados 11.362 ha de mata, 47.921 ha de lavouras anuais, 8.386 ha de pastagens, 3.669 ha de silvicultura e

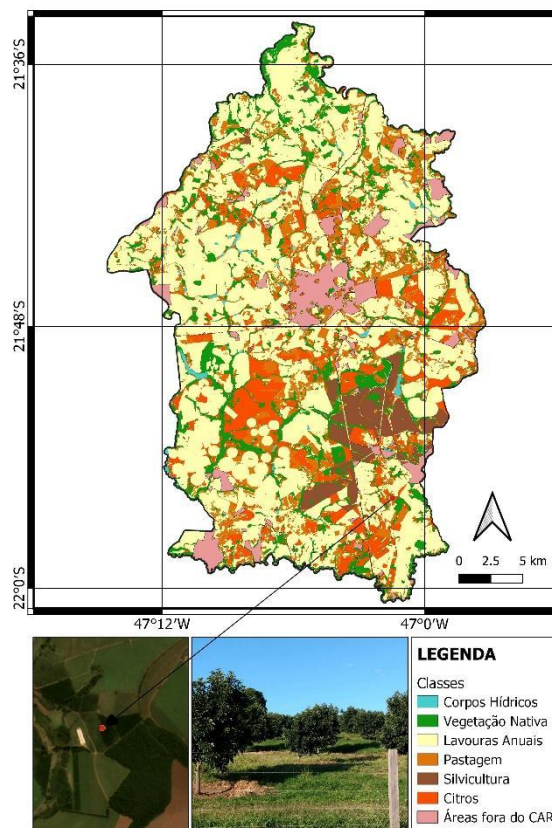


Figura 4. Mapa resultante da classificação de uso e cobertura da terra em 2021/2022, Casa Branca (SP).

No que diz respeito aos erros de classificação, houve maior confusão entre áreas de vegetação nativa, classificadas como citros, embora também tenha havido equívocos com áreas de pastagens e lavouras anuais, como mostrado na Tabela 2. Estudos prévios mostram que vegetação natural, áreas de pousio e de culturas perenes são comumente confundidas em classificações, e que o monitoramento do desenvolvimento da vegetação com séries temporais regulares, em diferentes anos, é um recurso ainda a ser explorado [18].

No Brasil, o projeto MapBiomass representa uma importante iniciativa de mapeamento de uso e cobertura da terra, em uma parceria multi-institucional que atualiza anualmente dados e metodologia. A continuidade das imagens Landsat e o potencial do processamento em nuvem na plataforma GEE permitiram o projeto gerar dados de 30 anos da dinâmica do uso e cobertura da terra. A partir da sexta coleção, lançada em 2021, o projeto incluiu pela primeira vez o mapeamento da citricultura, realizado por meio de classificações com mosaicos de medianas anuais e redes neurais convolucionais (U-Net). Especificamente para o ano de 2020, a metodologia estimou uma área de

aproximadamente 7.656 ha de citros para o município, o que representa uma diferença absoluta de 5.557 ha em relação aos dados da PAM. Assim como neste estudo, já foi avaliado que, apenas com uso de bandas espectrais do Sentinel-2, é possível separar áreas de citros de áreas sem citros com até 98% de precisão [9]. Ademais, informações acerca da elevação e declividade, o *Normalized Difference Vegetation Index* (NDVI) do Landsat 8, foram importantes para detectar citros em uma paisagem fragmentada com seis classes [8]. Dessa forma, em futuros estudos, as medianas mensais aqui utilizadas podem ser avaliadas em conjunto com modelos de elevação e textura, como estratégia para o contínuo aprimoramento da precisão no mapeamento da citricultura.

5. CONCLUSÕES

As medianas mensais de bandas espectrais dos satélites Sentinel-2 A e B, como variáveis de entrada no algoritmo de classificação digital *Random Forest*, são úteis para estimar a área e a distribuição espacial da produção de citros em região de elevada produtividade.

6. AGRADECIMENTOS

Os autores agradecem o apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo financiamento do projeto 2019/26222-6, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa do Programa Institucional de Bolsas de Iniciação Científica à primeira autora, e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de doutorado à segunda autora, Código Financeiro - 001.

REFERÊNCIAS

- [1] C. O. Almeida, O. S. Passos. *Citricultura brasileira em busca de novos rumos: Desafios e oportunidades na região Nordeste*. Embrapa Mandioca e Fruticultura, Cruz das Almas, 2011.
- [2] Associação Nacional dos Exportadores de Sucos Cítricos – CitrusBR. *A indústria brasileira de suco de laranja*. CitrusBR, São Paulo, 2011.
- [3] Instituto Brasileiro de Geografia e Estatística. Produção Agrícola Municipal. Disponível em <<https://sidra.ibge.gov.br/pesquisa/pam/tabelas>>. Acesso em: 20 out. 2022.
- [4] Fundação SEADE - Sistema Estadual de Análise de Dados. Perfil dos municípios paulistas. Disponível em <<https://perfil.seade.gov.br/>>. Acesso em: 19 out. 2022.
- [5] H. Xu, S. Qi, P. Gong, C. Liu, and J. Wang. Long-term monitoring of citrus orchard dynamics using time-series Landsat data: a case study in southern China. *International Journal of Remote Sensing*, v. 39, pp. 8271-8292, 2018.
- [6] A. Toosi, F. D. Javan, F. Samadzadegana, S. Mehravar, A. Kurban, and H. Azadi. Citrus orchard mapping in Juybar, Iran: Analysis of NDVI time series and feature fusion of multi-source satellite imageries. *Ecological Informatics*, v. 70, pp. 101733, 2022.
- [7] J. L. Hatfield., J. H. Pruejer, T. J. Sauer, C. Dold, P. O'Brien, K. Wacha. Applications of vegetative indices from Remote Sensing to agriculture: past and future. *Inventions*, v. 4, pp. 71, 2019.
- [8] F. L. Stefani. Zoneamento Geoambiental da Região de Casa Branca/SP. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2003.
- [9] C. A. Alvares, J. L. Stape, P. C. Sentelhas, J. L. M. Gonçalves, and G. Spavorek. Koeppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, v. 22, pp. 711-728, 2013.
- [10] H. N. Bendini, I. D. Sanches, T. S. Körtig, L. M. G. Fonseca, A. J. B. Luiz, and A. R. Formaggio. Using Landsat 8 image time series for crop mapping in a region of Cerrado, Brazil. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, v. XLI-B8, pp. 845-850, 2016.
- [11] H. P. S. Filho, A. F. J. Magalhães, and Y. S. Coelho. *Citros: O Produtor pergunta, a Embrapa responde*. Embrapa Informação Tecnológica, Brasília, 2005.
- [12] Serviço Florestal Brasileiro. Sistema Nacional de Cadastro Ambiental Rural - SICAR. Disponível em <<https://www.car.gov.br/publico/imoveis/index>>. Acesso em: 10 out. 2022.
- [13] L. Spinelli-Araujo; L. E. Vicente; C. V. Manzatto; L.A. Skorupa; D. de C. Victoria; A. R. Soares. Agrotag: um sistema de coleta, análise e compartilhamento de dados de campo para qualificação do uso e cobertura das terras no Brasil. In: XIX Simpósio Brasileiro de Sensoriamento Remoto, 2019. *Anais...* São José dos Campos: INPE, 2019.
- [14] J. C. D. Esquerdo, J. F. G. Antunes, A. C. Coutinho, E. A. Speranza, A. A. Kondo, and J. L. Santos. A web-based tool for visualization of MODIS vegetation indices in South America. *Computers and Electronics in Agriculture*, v. 175, pp. 105516.
- [15] Instituto Brasileiro de Geografia e Estatística. Produção da Extração Vegetal e da Silvicultura. <<https://sidra.ibge.gov.br/pesquisa/pevs/quadros/brasil/2021>>. Acesso em: 10 out. 2022.
- [16] G. Biau, and E. Scornet. A Random Forest guided tour. *Test*, v. 25, pp. 197-227, 2016.
- [17] L. Erpen., F. R. Muniz, T. S. Moraes, and E. C. R. Tavano. Análise do cultivo da laranja no Estado de São Paulo de 2001 a 2015. *Revista Pecege*, v. 4, pp. 33-43, 2018.
- [18] A. Bégué, D. Arvor, B. Bellon, J. Betbeder, D. Abelleira, R. P. D. Ferraz, V. Lebougeois, C. Lelong, M. Simões, S. R. Verón. Remote Sensing and cropping practices: a review. *Remote Sensing*, v. 10, pp. 99, 2018.