INTRODUÇÃO DE FERRAMENTAS DE BIG DATA E INTELIGÊNCIA ARTIFICIAL NA SELEÇÃO E PRÉ-PROCESSAMENTO DE IMAGENS ORBITAIS NO PROJETO PRODES

João Felipe S. K. C. Pinto[1], Igor P. Cunha[1], Jefferson J. de Souza[1], Lucélia S. de Barros[1], Gabriel C. Gonçalves[1], Bruna Maria P. Bento[1], Adeline M. Maciel[1], Daniel E Silva[2], Luciana S. Soler[1], Cassiano G. Messias[1], Vagner Luis Camilotti[1], Reynier Hernández[1], Ana Carolina S. de Andrade[1], Douglas Rafael Vidal de Moraes[1], Luis E. Maurano[1], Cláudio A. de Almeida[1]

[1]National Institute for Space Research (INPE), São José dos Campos, 12227-010, Brazil; [2]World Wide Fund for Nature (WWF), Brasília, 70377-540 Brazil;
First author's e-mail: jfkneipp@gmail.com

RESUMO

O PRODES monitora a cobertura da vegetação primária nos biomas Amazônia e Cerrado utilizando imagens de satélite de 353 órbitas-ponto Landsat. Desse total, 227 estão na Amazônia Legal e 126 no Cerrado. O número final de imagens para cobrir essas órbitas-ponto pode variar em função da cobertura de nuvens, levando ao uso de imagens auxiliares. Em 2021, ambos os biomas foram cobertos por 793 imagens, totalizando 3172 etapas de processamento digital de imagens (PDI), além de etapas complementares (detecção de nuvens e geração de mosaicos). Tecnologias de inteligência artificial unidas ao acesso a imagens de satélite em plataformas de *big data* e o processamento em nuvem têm permitido a otimização e automatização do PDI, reduzindo o tempo de mapeamento. Neste estudo, demonstramos que a aplicação dessas ferramentas de automatização apontam para uma otimização de 71% e 25%, respectivamente, do tempo necessário para seleção/download/composição e PDIs complementares para o efetivo mapeamento.

Palavras-chave — PRODES, *big data*, automatização, AI, imagens orbitais.

ABSTRACT

PRODES monitors primary vegetation cover in the Amazon and Cerrado biomes using satellite images from 353 Landsat path and rows. Of this total, 227 are in the Legal Amazon and 126 in the Cerrado. The total amount of images to cover the paths and rows may vary according to cloud cover, leading to the use of auxiliary images. In 2021, both biomes were covered by 793 images, totaling 3172 digital image processing (DIP) steps, in addition to complementary steps (cloud detection and mosaic generation). Artificial intelligence technologies combined with access to satellite images on big data platforms and cloud processing have allowed DIP optimization and automatization, reducing mapping time. In this study, we demonstrate that the application of these automation tools leads to an optimization of 71% and 25%, respectively, of the time required for selection/download/composition and complementary DIPs for effective mapping.

Keywords — *PRODES, big data, automation, AI, orbital images.*

1. INTRODUÇÃO

O Programa de Monitoramento da Floresta Amazônica Brasileira por Satélite (PRODES), criado pelo Instituto Nacional de Pesquisas Espaciais (INPE) em 1988, passa continuamente por um processo de aperfeiçoamento de suas técnicas analíticas. Na fase inicial do projeto, entre 1988 e 2002, conhecida como PRODES analógico, o desmatamento era identificado de maneira visual sobre composições coloridas do sensor Thematic Mapper (TM) (Landsat 5), impressas na escala 1:250.000. Os novos incrementos eram delimitados em papel vegetal sobreposto às imagens e posteriormente digitalizados. Considerando os custos e dificuldades logísticas, a totalidade das imagens utilizadas não recobria toda a Amazônia Legal Brasileira [1].

O avanço tecnológico e o desenvolvimento da capacidade técnica e computacional do INPE proporcionaram a transição do PRODES analógico para o digital a partir de 2000. Esta fase foi marcada pela utilização de algoritmos semi-automatizados de segmentação, classificação, pós-classificação e ferramentas de edição matricial e vetorial dentro de um banco de dados georreferenciado que manteve total coerência com os dados já produzidos

Com o advento do sistema de informação geográfica (SIG) TerraAmazon em 2005, permitindo o gerenciamento de banco de dados e edição vetorial multiusuário, foi adotada a fotointerpretação realizada diretamente na tela do computador, permitindo uma maior acurácia na classificação [2]. Imagens obtidas pelos satélites Sentinel 2B/2A e CBERS 4/4A passaram a ser utilizadas como auxiliares sempre que as imagens Landsat estivessem indisponíveis no período ideal e livre de nuvens nas áreas de interesse para o mapeamento, reduzindo significativamente a área não observada a partir de 2019 [3].

Em 2012, o PRODES passou a monitorar o bioma Cerrado que, conjuntamente com o monitoramento da Amazônia, passou a exigir o processamento (p.ex., seleção, composição, criação de mosaico, disponibilização das imagens, detecção de nuvens) de 353 imagens primárias, número que pode aumentar em função da disponibilidade de imagens com baixa cobertura por nuvens. Com a cobertura dos biomas Caatinga, Pampa, Pantanal e Mata Atlântica a partir de 2023, a demanda processual do projeto será consideravelmente ampliada e, embora até o ano de 2022 o PRODES continue utilizando fotointerpretação, inúmeras etapas de automatização vêm sendo implantadas. Um relevante exemplo disso é a detecção automática de nuvens que visa reduzir o tempo de interpretação sem alterar a qualidade e confiabilidade do dado PRODES.

Assim, com o intuito de publicizar a contínua evolução do sistema PRODES, este trabalho descreve e analisa os ganhos da introdução de ferramentas de *big data*, bem como de inteligência artificial para automatização de processos no projeto PRODES nos biomas Amazônia e Cerrado.

2. MATERIAL E MÉTODOS

Para a realização da análise dos ganhos associados à introdução de novas tecnologias, inicialmente foram selecionadas e contabilizadas todas as etapas, as tarefas de processamento digital de imagens (PDI) e as órbitas-ponto que compõem o projeto PRODES para os biomas Amazônia e Cerrado. Foram realizados testes de viabilidade de implementação de processos de automatização durante pelo menos 6 meses, que incluíram inicialmente o desenvolvimento de códigos e, em um segundo momento, o treinamento de usuários finais (intérpretes) na utilização das novas rotinas de produção. Os códigos foram produzidos utilizando a plataforma de processamento em nuvem Google Earth Engine (GEE), de automação GitHub, AWS e de *bots* do aplicativo de voz sobre IP, Discord. Os testes foram realizados sobre as seguintes tarefas essenciais de PDI: 1) Seleção e composição de imagens da família Landsat e Sentinel; 2) Download e disponibilização de imagens aos intérpretes; 3) Criação de mosaicos Sentinel e 4) criação de camada vetorial delimitando áreas cobertas por nuvens e suas sombras.

Para mensurar o ganho de tempo da implementação destes procedimentos automatizados, foi selecionada a etapa de maior consumo de tempo para o projeto: a seleção de imagens dentro do período ideal para detecção do desmatamento com a menor cobertura possível de nuvens para cada órbita-ponto, conforme metodologia PRODES publicada [4]. Assim, a partir da comparação entre o tempo gasto na seleção semi-automatizada utilizando o GEE e o tempo de elaboração pelo método tradicional, conforme descrito na metodologia atual do projeto, foi possível quantificar a melhoria dos processos e avaliar os ganhos de otimização. A comparação ocorreu nas duas órbitas-ponto de maior quantidade de polígonos de desmatamento identificados no ano de 2021. O uso de diferentes imagens de equivalente grau de dificuldade fez-se necessário para eliminar um eventual ganho de tempo pela familiaridade do intérprete com a imagem.

3. RESULTADOS E DISCUSSÃO

3.1 Seleção e composição de imagens da família Landsat e Sentinel

Com o intuito de potencializar e automatizar a seleção de imagens utilizadas no mapeamento do PRODES, todo o processo passou a ser testado em ambiente virtual por meio do GEE, através de códigos de programação que transferem as etapas, antes executadas em mais de um ambiente, para dentro da plataforma (Figura 1). O processo atual, realizado de forma manual, contempla as etapas seguintes: acessar o servidor virtual de imagens; selecionar uma cena; transferir o arquivo para um armazenamento local; adicionar em um SIG; processar as imagens; e sobrepor arquivos vetoriais que auxiliam a tomada de decisão. Esse processo passaria a ser executado em nuvem, dentro do servidor do GEE, proporcionando um ganho expressivo de tempo, uma vez que os resultados são exibidos em poucos segundos. No GEE é também adicionada a camada vetorial referente à máscara de exclusão que contém a área desmatada acumulada até o ano anterior, o que permite selecionar a melhor imagem considerando apenas a área que deverá ser observada no ano corrente.

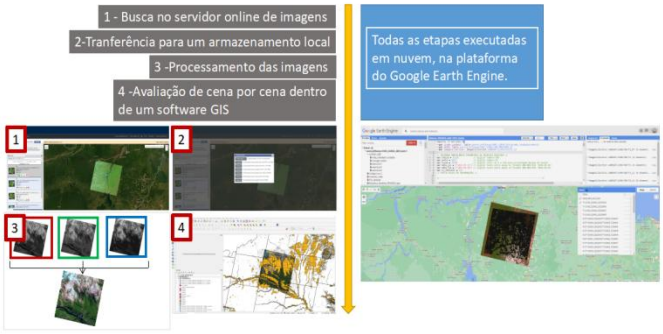


Figura 1. Comparação entre os métodos de seleção: À esquerda da seta amarela, o método anterior, com as etapas indicadas pelos números. À direita, a tela de exibição dos resultados no GEE.

3.2 Disponibilização de imagens para download aos intérpretes

Após o processo de seleção das melhores cenas para mapeamento do PRODES, é realizado o download das imagens. Na maneira tradicional, são realizadas várias etapas até o download da imagem, tais como: fazer login na interface de acesso ao servidor, percorrer o diretório de

interesse, realizar a busca da imagem e download. Para este processo, está sendo desenvolvido um conjunto de *scripts* em linguagem Python, usando o *Application Programming Interface* (API) Request e SentinelSat, que possibilitam acesso ao *United States Geological Survey* (USGS) para download das imagens Landsat e acesso ao Copernicus *Open Access Hub* para imagens Sentinel. As composições RGB das bandas 5/4/3 (Landsat 5 e 7), bandas 6/5/4 (Landsat 8 e 9) e bandas B11, B8A, B04 (Sentinel) são geradas automaticamente por meio do *script* e ficam armazenadas em um diretório do servidor do INPE que, posteriormente, será acessado pela equipe de intérpretes. Esses farão então o contraste de cada imagem. Visando automatizar e otimizar o acesso às imagens armazenadas, foi desenvolvido um *bot* (do inglês, robô) utilizando a API própria para o Python para o aplicativo gratuito de comunicação Discord [5], ferramenta utilizada pela equipe do projeto para comunicação interna. *Bots* permitem aos intérpretes, a partir de simples comandos de texto, acesso imediato ao link para *download* das imagens a partir do seu identificador (ID). A Figura 2 demonstra o uso dos comandos $id e $ctrs que retornam, respectivamente, o link de uma imagem Landsat 9 com composição colorida e os links de acesso às imagens Sentinel com contraste aplicado.

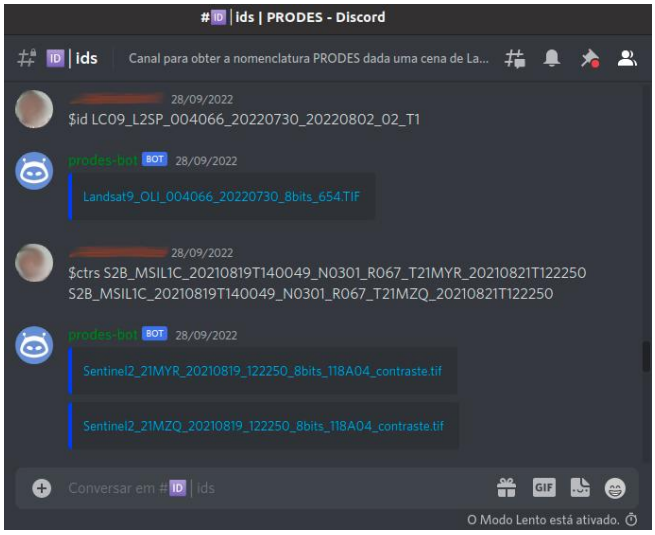


Figura 2. Exemplo de uso dos comandos $id e $ctrs no Discord

3.3 Criação de mosaicos Sentinel

O mapeamento PRODES é realizado, historicamente, a partir das órbitas/ponto Landsat. Porém, a disponibilização de novos sensores com características semelhantes, nos últimos anos representa uma oportunidade para reduzir o tempo de revisita. A ausência de imagens Landsat com baixa cobertura de nuvem, pode ser compensada por imagens Sentinel, disponibilizados para os intérpretes na forma de mosaicos de imagens. Para isto, está sendo testado um script na plataforma GEE que permite realizar a seleção da imagem Sentinel com menor percentual de nuvens e dentro do período ideal da cena (ver metodologia PRODES em Almeida et al., 2022). A rotina do *script* executa a função mosaico, que une diversos *tiles* da Sentinel 2A e gera um produto único em composição colorida RGB. O mosaico é recortado pela área de interesse da órbita/ponto Landsat, originando, assim, a imagem que será utilizada no mapeamento (Figura 3). A rotina executada pelo código também exporta um arquivo vetorial que contém a data, ID e percentual de nuvens de cada tile do mosaico, que será usado pelos intérpretes para mapear os polígonos de desmatamento com a data original de detecção. Os arquivos são direcionados para o Google Drive e, depois, armazenados no servidor do INPE.

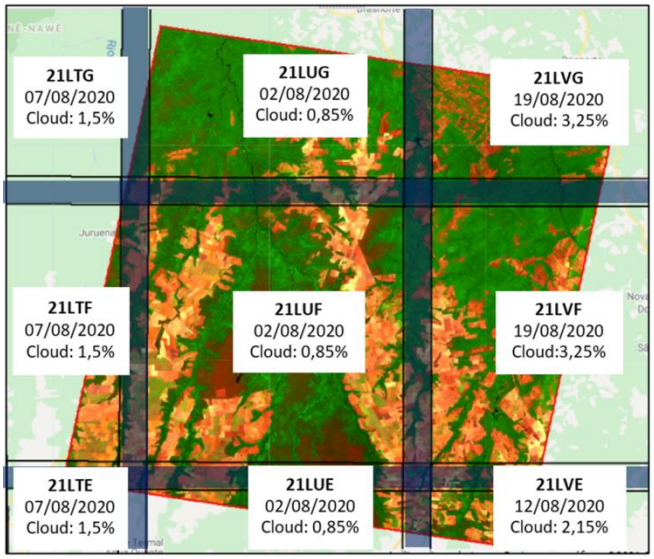


Figura 3. Relação de sobreposição de tiles Sentinel-2, na Cena 228/69. Contorno em preto: Grade Sentinel; Contorno em vermelho: Grade Landsat; Retângulos translúcidos em azul escuro: Sobreposição de tiles.

3.4 Criação de camada vetorial sobre áreas de nuvens

A metodologia desenvolvida para detecção de nuvens utiliza diferentes ferramentas de PDI, como o modelo linear de mistura espectral, índices normalizados e métricas de textura (GLCM). Um total de 224 variáveis compõem os dados de entrada para o algoritmo de classificação utilizado, o *Random Forest* (Figura 4). A metodologia se demonstrou eficaz na detecção de nuvens, sendo geradas as seguintes classes: nuvens opacas (nuvens em que não é possível observar a cobertura do solo); nuvens translúcidas (nuvens em que é possível observar a cobertura do solo e/ou possuem mistura espectral); sombra de nuvem (classe que representa a sombra das nuvens).

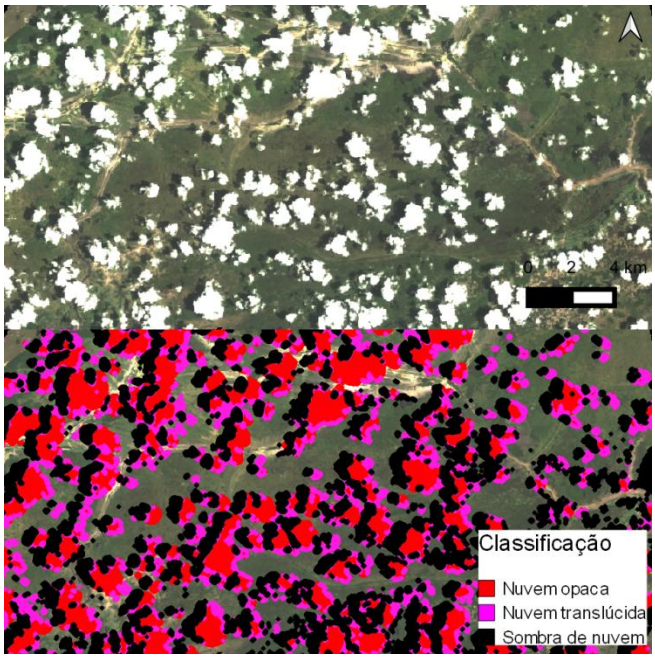


Figura 4. Resultado da classificação de nuvens pelo algoritmo *Random Forest*, com três classes de nuvens.

3.5 Avaliação de tempo na seleção e composição de imagens Landsat e Sentinel

Para Visando avaliar o ganho de tempo direto da implementação, foram selecionadas as órbitas- ponto com maiores valores percentuais de desmatamento para o ano de 2021 (Tabela 1).

Os resultados preliminares apontam a viabilidade na adoção do método automatizado e sua expansão para todo o território monitorado no projeto PRODES para o processo de seleção e composição de imagens. Além disso, esses resultados mostram uma economia média de 71,6% no tempo necessário para conclusão das quatro etapas do processo, otimizando o processamento digital de imagens.

Órbita Ponto	% de desmatamento sm 2021	Tempo para processamento manual
232/66	4,13%	19’12’’
227/65	3.82%	18’56’’

Tabela 1. Comparação entre tempo necessário para a realização do processo manual e automatizado de seleção e composição de imagens Landsat e Sentinel

Com base na quantidade de imagens utilizadas para mapeamento dos biomas Amazônia e Cerrado para o ano de 2021, um total de 793 imagens, a adoção deste processamento, somente para esta etapa, refletiria em uma economia de 178,4 horas de trabalho, permitindo a alocação de um ou mais profissionais por um menor período e por consequência permitiria a utilização deste tempo em outras atividades componentes do projeto.

5. CONCLUSÕES

Todos os processos de automatização aplicados pela plataforma GEE se mostraram eficazes na eliminação de etapas, assim como mostraram um ganho significativo de tempo. Embora ainda seja necessário um estudo de viabilidade de implementação, as etapas de processamento digital de imagens apontados neste documento indicam um cenário positivo na futura adoção de etapas automatizadas em substituição das atuais etapas manuais e individualizadas. Uma eventual adoção destas etapas automatizadas, através de testes, também permitiria uma mais fácil adição de etapas de avaliação do dado, tais como auditorias ou validações, que juntos proporcionariam uma maior confiabilidade no dado gerado, eliminando uma eventual contaminação do dado devido a subjetividade gerada somente pela experiência de um profissional.

8. REFERÊNCIAS

[1] INPE. *Desflorestamento 1995-1997 Amazônia*. São José dos Campos: INPE, 1997. Disponível em: <http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes/pdfs/prodes1995-1997.pdf>. Acesso em: 13 mar. 2021.

[2] L. E. P. Maurano. *Avaliação da qualidade dos dados PRODES: estimativa e regionalização dos erros de mapeamento*. Dissertação (mestrado) - Pós-Graduação em Sensoriamento Remoto, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2018.

[3] L. Soler,, D. Silva, C. G. Messias et al. Promising advances of Amazon monitoring systems throughout vanguard technology and scientific knowledge, The Intern. *Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, v. XLIII-B3-2021, 2021.

[4] C. A. Almeida et al. *Metodologia utilizada nos sistemas Prodes e DETER - 2a edição (atualizada)*. Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2022. Disponível online: <http://urlib.net/8JMKD3MGP3W34T/47GAF6S>

[5] A. Verma, S. Tyagi, G. Mathur. A Comprehensive Review on Bot-Discord Bot. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, p. 532-536, 2021.