

# RSDD: um Conjunto de Dados para Modelagem de Desmatamentos

Rosangela Maria Guarienti Ventura<sup>1,2</sup>, Carlo Ralph de Musis<sup>1,2</sup>, Thiago Meirelles Ventura<sup>1</sup>, Adriano Marcelo Vilarga Corbelino II<sup>1</sup>

<sup>1</sup>Universidade Federal de Mato Grosso (UFMT)  
Cuiabá, Mato Grosso – Brasil

<sup>2</sup>Perícia Oficial e Identificação Técnica (POLITEC)  
Cuiabá, Mato Grosso – Brasil

rosangelaventura@politec.mt.gov.br, carlo.demusis@gmail.com,  
thiago@ic.ufmt.br, adriancorbelinoii@gmail.com

**Abstract.** *It is common to use remote sensing to understand the dynamics of deforestation. However, obtaining and processing these images is not a trivial task, besides to the difficulty of having validated deforestation areas. To solve this problem, this work presents the Remote Sensing Deforestation Dataset (RSDD) containing tabulated data and images of areas affected by deforestation validated by a state agency. RSDD has a size less than 1% related to the original remote sensing images, provides a reduction in efforts in future research and can serve as a common dataset to be used as benchmark for machine learning solutions.*

**Resumo.** *É comum o uso de sensoriamento remoto para compreender a dinâmica de desmatamentos. Entretanto, a obtenção e processamento dessas imagens não é uma tarefa trivial, além de haver a dificuldade de ter áreas validadas de desmatamento. Para resolver esse problema este trabalho apresenta o Remote Sensing Deforestation Dataset (RSDD) que contém dados tabulados e imagens de áreas afetadas pelo desmatamento validado por um órgão Estadual. O RSDD possui um tamanho menor que 1% em relação às imagens originais de sensoriamento remoto, proporciona a redução de esforços em futuras pesquisas e pode servir como uma fonte comum para ser utilizada como benchmark em soluções de aprendizado de máquina.*

## 1. Introdução

Um recurso para combater desmatamentos é o sensoriamento remoto. Trabalhos como os de Maurano et al. (2019) utilizam desse recurso para realizar estimativas e monitoramento de áreas desmatadas. Entretanto, há dificuldades em relação a esse tipo de trabalho, sendo uma delas à respeito das informações de áreas de desmatamento. São necessárias diversas ocorrências de desmatamento para ser possível realizar alguma análise ou treinamento de métodos estatísticos ou de aprendizado de máquina. Outra dificuldade é que cada imagem gerada por um satélite consome uma grande quantidade de espaço, dificultando a obtenção e armazenamento desse tipo de dado.

Para contornar essas dificuldades, este trabalho disponibiliza o Remote Sensing Deforestation Dataset (RSDD), um conjunto de dados elaborado para análise e

modelagem de desmatamentos, no qual os dados são provenientes de imagens de satélite e validados com áreas desmatadas ilegalmente. O RSDD está em um formato mais leve e compatível com as principais bibliotecas de modelagem de dados.

## **2. Materiais e Métodos**

Os dados do RSDD são referentes ao Estado de Mato Grosso (MT), que é um dos nove estados da Amazônia Legal e apresenta alto índice de desmatamento. Em Dias et al. (2018) foram reunidas análises dos desmatamentos ilegais ocorridos no ano de 2017 em MT, validadas por analistas ambientais da Secretaria de Estado do Meio Ambiente. Nesse relatório, os dados de desmates de 2016 foram utilizados para verificar a diferença de cobertura vegetal em 2017, possibilitando a seleção de apenas novos desmatamentos, ou seja, que ocorreram em 2017.

Além das informações referentes às áreas desmatadas em MT, foram utilizados dados de satélite de cada área. Os dados foram provenientes dos satélites da missão Copernicus Sentinel-2 (ESA, 2021). Baseado nas coordenadas das áreas desmatadas, foram obtidas as imagens do Sentinel-2 que continham os pontos selecionados. As imagens foram analisadas descartando pontos que tivessem valores nulos em suas bandas. O mesmo foi feito se houvesse a presença de nuvens ou corpos d'água, utilizando o algoritmo Fmask (ZHU et al., 2015). Foi também definido o tamanho mínimo da área de interesse (AOI) no valor de 3.600 m<sup>2</sup>.

Para a manipulação das imagens de satélite foi utilizada a linguagem de programação Python em conjunto com as bibliotecas GDAL e Rasterio. Em seguida, foram gerados arquivos com os dados tabulados no formato CSV, assim como as imagens que representam as AOIs em cada imagem de satélite. As imagens foram extraídas da True Color Image (TCI) que acompanha as bandas da imagem, sendo feita uma normalização dos valores das bandas (RGB), possibilitando uma melhor visualização.

Os dados tabulados em arquivos texto no formato CSV são úteis para aplicação de métodos de aprendizagem de máquina, no qual os dados já estão em formato numérico e estruturados. Já as imagens, estão no formato PNG e foram centralizadas no ponto da coordenada que houve o desmate, podendo servir para verificação dos resultados e análises mais detalhadas.

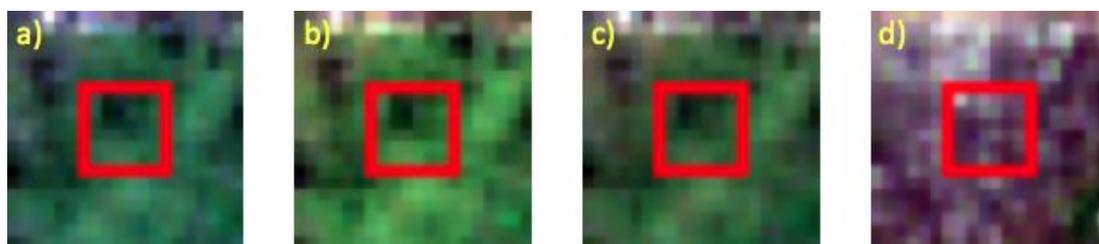
É importante saber quando aconteceu a transição de uma área conservada para uma área desmatada para que os métodos possam analisar as diferenças e, assim, aprender quando uma área está próxima de ser desmatada. A abordagem aplicada neste trabalho seguiu a metodologia de Bayma & Sano (2015), no qual compararam índices de vegetação para definir se um local estava desmatado ou não no bioma do Cerrado, determinando limiares dependendo da formação vegetacional. O NDVI obteve os melhores resultados, então foi feita uma análise com esse índice para determinar limiares de desmatamento de acordo com as áreas de MT, que foi dividida em 25 quadrantes. Para cada quadrante foi realizada uma análise visual das imagens existentes a fim de anotar com quais valores de NDVI aconteciam os desmatamentos. Como os valores variavam, foram definidos limiares inferiores e superiores manualmente para cada quadrante, sendo que valores de NDVI menores que o limiar inferior significam que a área está desmatada e valores maiores que o limiar superior significam que a área estava conservada. Foram evitadas áreas com NDVI entre o limiar inferior e superior

por ter um menor grau de confiança na classificação. Com esses limiares foi possível classificar cada imagem obtida, sendo ela desmatada ou conservada, tornando possível a elaboração do conjunto de dados RSDD.

### 3. Conjunto de Dados RSDD

Os dados tabulados estão divididos em três arquivos CSV. O primeiro deles, chamado “points.csv” contém informações sobre os pontos existentes no conjunto de dados, com seu ID, latitude, longitude e os hectares desmatados. O segundo arquivo é o “bands.csv”, no qual possui dados extraídos das bandas das imagens de satélite para cada ponto do conjunto de dados, com ID do ponto, data, índice do ponto e os valores das 13 bandas do Sentinel-2. Por fim, o arquivo “imgs.csv” tem informações sobre as imagens geradas, contendo o ID do ponto, data e a URL da imagem original em que foi extraída os dados dos pontos. Com esses dados é possível treinar métodos de aprendizado de máquina para realizar as classificações de desmatamento.

Para a organização das imagens, o nome dos arquivos foi padronizado, os quais contém informações sobre o ID do ponto e data da imagem. Em conjunto com as informações armazenadas em “point.csv” e “imgs.csv”, o pesquisador possui todos os dados necessários para processar tais imagens e, assim, auxiliar em suas aplicações. Há mais de uma imagem para cada ponto, representando a evolução do desmatamento. Na Figura 1 são apresentados quatro momentos representando áreas que parecem estar conservadas e uma área totalmente desmatada. Ao todo, 6.400 imagens estão disponibilizadas neste conjunto de dados. Cada imagem cobre uma área de 60x60m de locais onde houve desmate ilegal. As imagens estão no formato PNG com resolução 20x20 pixels, o que abrange a área representada do desmate e o seu entorno.



**Figura 1. Exemplo de imagens de um único ponto do RSDD, no qual em a), b) e c) a área parece estar conservada, mas em d) a área já está totalmente desmatada. A diferença de tempo entre c) e d) foi de menos de 30 dias. A área retangular vermelha consiste em um espaço de 60mx60m, que é a menor resolução espacial do Sentinel-2.**

Uma vantagem do RSDD é a facilidade em obter dados essenciais para uma modelagem de desmatamento, tanto em relação à leveza quanto ao formato dos dados. Os dados originais somam 267 GB de imagens do sensoriamento remoto, enquanto no RSDD há apenas 79,3 MB. Essa diferença se dá principalmente pelo pré-processamento realizado, deixando somente informações válidas e pertinentes às áreas de desmatamento. As imagens do satélite são disponibilizadas em grânulos de tamanho de 100km por 100km, ocupando cerca de 700MB por imagem. Neste tipo de pesquisa não é necessário utilizar toda a área que uma imagem pode abranger, mas sim somente as áreas de interesse.

Além da diferença de espaço para download, armazenamento e processamento, há também a vantagem deste conjunto de dados disponibilizar o mapeamento e

validação de várias áreas de desmate. Isso é algo necessário para o treinamento de métodos estatísticos ou de aprendizado de máquina. Neste conjunto de dados estão agrupados 800 pontos de área de desmate ilegal validadas. Para cada ponto existe a imagem do momento do desmate e três imagens de momentos anteriores ao fato, dentro de um espaço temporal de três meses. Além disso, outros 800 pontos de áreas conservadas foram adicionados no RSDD para que seja possível fazer comparações com as áreas desmatadas.

#### **4. Considerações Finais**

Este trabalho apresentou um novo conjunto de dados, denominado Remote Sensing Deforestation Dataset (RSDD), que contém dados de áreas desmatadas ilegalmente da Amazônia Legal. O RSDD visa facilitar as pesquisas que envolvem desmatamento, fornecendo de maneira compacta os dados necessários para modelagem deste tipo de estudo. Os dados estão em dois formatos: tabulados em arquivos CSV com os valores extraídos das bandas e em imagens PNG com resolução 20x20. Assim, os pesquisadores não precisam validar dados de desmate ou executar procedimentos de obtenção e extração em imagens de satélite. Os dados fornecidos são suficientes para processá-los com métodos estatísticos ou de aprendizado de máquina, incluindo deep learning com o uso das imagens. O conjunto de dados RSDD e os códigos relacionados estão disponibilizados no repositório <https://github.com/demusis/srdesmatamentos>.

#### **Agradecimentos**

Os autores agradecem a Secretaria de Estado do Meio Ambiente (SEMA) de Mato Grosso, o portal Earth Explorer e a missão Copernicus Sentinel-2 pelo fornecimento dos dados utilizados neste trabalho.

#### **Referências**

- Bayma, A. P. and Sano, E. E. (2015). Séries temporais de índices de vegetação (ndvi e evi) do sensor modis para detecção de desmatamentos no bioma cerrado, *Boletim de Ciências Geodésicas* 21(4): 797–813.
- Dias, A., Silva, K. L., Souza, L. G., Perdigão, R. J., Bilio, H., C., O. H. H. and Kummer, O. P. (2018). Análise dos dados de desmatamento do estado de mato grosso: Período 2016/2017, Relatório Técnico 003/2018/CGMA/SRMA/SAGA/SEMA-M, Secretaria de Estado do Meio Ambiente de Mato Grosso (SEMA-MT), Mato Grosso, Brasil.
- ESA (2023). Sentinel-2 Mission. Available at <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>.
- Maurano, L., Escada, M. and Renno, C. (2019). Padrões espaciais de desmatamento e a estimativa da exatidão dos mapas do prodes para amazônia legal brasileira, *Ciência Florestal* 29(4): 1763–1775.
- Zhu, Z., Wang, S. and Woodcock, C. E. (2015). Improvement and expansion of the fmask algorithm: cloud, cloud shadow, and snow detection for landsats 4–7, 8, and sentinel 2 images, *Remote Sensing of Environment* 159: 269–277.