

# Extração de Séries Espaço-Temporais a partir de um Cubo de Dados Geoespacial em Benefício da Cafeicultura Mineira

Marcelo Robert Santos<sup>1</sup>, Melise M. V. Paula<sup>1</sup>, Vanessa C. O. Souza<sup>1</sup>

<sup>1</sup>Instituto de Matemática e Computação – Universidade Federal de Itajubá  
(IMC-UNIFEI) – Caixa Postal 50 – 37500-903 – Itajubá – MG

(marcelorobertsantos, melise, vanessasouza)@unifei.edu.br

**Abstract.** *Coffee production is a pre-eminent and significant agricultural activity in the Brazilian economy. That said, this article seeks to explore a model for storing and processing vegetation indices for coffee growing obtained through orbital images from the Sentinel-2 satellite stored in the Brazil Data Cube. The data flow and data structure were defined, having selected the document-based database MongoDB due to the spatial and temporal query functions.*

**Resumo.** *A cafeicultura é uma das principais atividades agrícolas do Brasil. Dada sua importância para a economia brasileira, o presente artigo busca explorar um modelo de armazenamento e processamento de índices de vegetação voltados para a cafeicultura obtidos através de imagens orbitais do satélite Sentinel-2 armazenados no Brazil Data Cube. O fluxo e a estrutura dos dados foram definidos, tendo selecionado o banco de dados baseado em documentos MongoDB devido a funções de consulta espaciais e temporais.*

## 1. Introdução

O café representa uma das principais *commodities* do país. A tradição da cafeicultura em Minas Gerais desempenha um papel crucial na produção nacional desse grão. Segundo [Vieira et al. 2010], o monitoramento da cultura é imprescindível para uma cafeicultura mais sustentável e competitiva. O sensoriamento remoto em conjunto com a geoinformática oferece agilidade e precisão no levantamento de informações ambientais, propiciando uma melhora no planejamento e tomada de decisões.

Esse monitoramento pode ser melhorado com o uso dos índices de vegetação (IVs) gerados a partir das imagens de satélite. Isso porque os IVs estão correlacionados aos parâmetros agronômicos de uma cultura, como estresse hídrico, nutrientes no solo e na planta e perda foliar [Souza et al. 2019]. Alguns dos índices frequentemente utilizados na cafeicultura incluem o NDVI (*Normalized Difference Vegetation Index*); o EVI2 (*Two Band Enhanced Vegetation Index*); e o NDWI (*Normalized difference Water Index*).

Apesar dos avanços na tecnologia de sensoriamento remoto, atualmente existe uma demanda por aplicações que facilitem o acesso a esses dados. Recentemente o Brasil passou a contar com o projeto *Brazil Data Cube - BDC*<sup>1</sup>, que fornece uma infraestrutura de dados georreferenciados (imagens de satélite) em formato de cubo de dados multi-dimensionais e ferramentas e *softwares* para sua manipulação. Ao manipular o cubo, é possível extrair informações de uma determinada região, tanto para uma única data,

---

<sup>1</sup><http://www.brazildatacube.org/>

quanto uma série temporal. Contudo, ainda há um desafio quanto à complexidade técnica e curva de aprendizado para a utilização de tais ferramentas.

O objetivo desse trabalho é relatar a experiência em utilizar o BDC para gerar uma base de dados formada por dados espectrais e diferentes índices de vegetação com intuito de potencializar pesquisas para a cafeicultura do Sul de Minas. A proposta é que essa base de dados seja utilizada no desenvolvimento de diferentes aplicações para auxiliar os atores da cadeia produtiva do café na tomada de decisão.

Ao integrar tecnologias de sensoriamento remoto em práticas agrícolas modernas, é possível alcançar um equilíbrio entre a produção agrícola e a conservação dos recursos naturais. Nesse sentido, esse trabalho torna-se relevante ao propiciar acesso facilitado de dados de sensoriamento remoto à cadeia produtiva do café. Esse trabalho está em andamento e compõe um projeto maior que objetiva democratizar informações sobre a cafeicultura no Estado de Minas Gerais.

## 2. Materiais e Métodos

Este trabalho tem natureza aplicada, com abordagem qualitativa e caráter descritivo, buscando estudar uma forma de armazenamento e disponibilização de séries temporais de dados georreferenciados sobre o café. Sendo assim, esse trabalho explora os dados do BDC a fim de gerar uma base de dados própria, que subsidiará aplicações voltadas para atores que compõem a cadeia produtiva cafeeira de Minas Gerais. Como projeto piloto, o município mineiro de Três Pontas foi escolhido, tendo em vista sua importância para a cafeicultura em Minas Gerais.

As imagens utilizadas são da coleção S2-16D-2 do satélite Sentinel 2. Para a base que está sendo construída, foram selecionados para armazenamento as bandas espectrais<sup>2</sup> do azul, verde, vermelho e infravermelho próximo, além dos índices espectrais NDVI, EVI2 e NDWI. O fluxo de trabalho é detalhado na Figura 1. As etapas estão sendo desenvolvidas na linguagem R, versão 4.3.1.

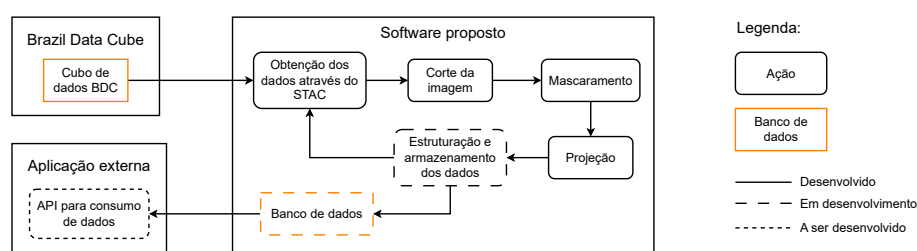


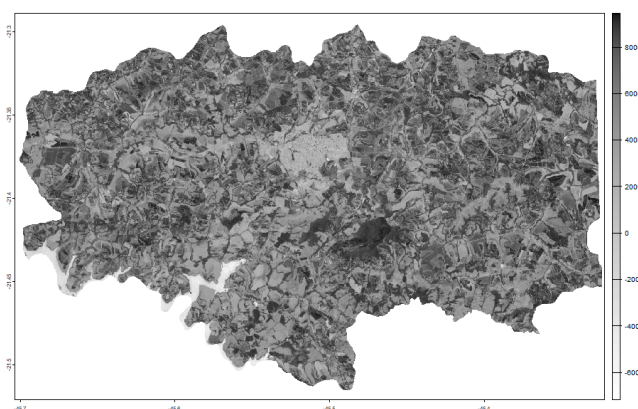
Figura 1. Fluxo de dados

O tratamento dos dados é feito de forma sequencial, iniciando ao obter as bandas desejadas através de suas urls, filtrando-as por data e localização. Após obter a imagem, a mesma é recortada para a extensão do município de Três Pontas e é gerada uma máscara para que *pixels* fora da região desejada tenham seus valores desconsiderados. A próxima etapa é de reprojeção cartográfica. Os dados do BDC estão na projeção Albers Equal-Area Conic, utilizada para grandes extensões territoriais, como países e continentes. No caso

<sup>2</sup>Bandas espectrais referem-se a faixas do espectro eletromagnético. O satélite coleta informações em diferentes bandas e, cada banda, gera uma imagem.

desse trabalho, foi utilizado o sistema de coordenadas de latitude e longitude, com Datum WGS84 (EPSG:4326), o qual é compatível com o GeoJSON utilizado no MongoDB.

Esse processo é repetido para todas as datas de passagem do satélite Sentinel 2 e, para cada data, a informação de cada pixel da área de estudo é armazenada em um banco de dados geográfico. O SGBD escolhido foi o MongoDB devido ao fato de ele ser nativamente espacial e, por ser orientado a documentos e flexibilizar a modelagem. Ademais, o MongoDB possui um tipo de coleção chamada *Time Series Collection*<sup>3</sup> que armazena eficientemente séries temporais, melhorando a eficiência das consultas e reduzindo o uso de disco. Um exemplo do resultado do processamento pode ser visto na Figura 2.



**Figura 2. Dados obtidos a partir do BDC. NDVI - 10/06/2022.**

### 3. Resultados

Esse trabalho atualmente encontra-se na fase de definição da modelagem utilizada na base de dados. Foram deslumbradas duas modelagens com agregados (unidade conceitual representada) diferentes. Na primeira, cada documento seria um *pixel*. E, para cada *pixel*, além de sua localização, estariam associados *arrays* dos atributos (bandas espectrais e índices de vegetação). Os *arrays* são compostos pelo par data e valor. A outra modelagem é aquela que utiliza a coleção *time series* do MongoDB. O documento representa *um pixel em uma data* e contém o valor de todos os índices. Um exemplo de ambas as modelagens pode ser visto na Figura 3.

Fica claro que na modelagem da Figura 3a, são gerados menos documentos (um para cada *pixel*). Porém, os documentos podem ficar grandes, uma vez que a coleção S2-16D-2 do BDC possui uma resolução temporal de aproximadamente 16 dias. Com isso, a cada 16 dias haveria uma atualização nesses documentos, incluindo os valores coletados na data avaliada. Ressalta-se também, no entanto, que o dado em si não é grande, sendo composto apenas por uma data e um valor numérico.

Já na modelagem utilizando a coleção *time series* (Figura 3b), são gerados diversos documentos, sendo um para cada *pixel* por data. Nesse caso, os documentos são pequenos e, a cada nova passagem do satélite, novos documentos são criados e não atualizados. Nessa modelagem há de se considerar também a estrutura da coleção, que é otimizada para lidar com séries temporais.

<sup>3</sup><https://www.mongodb.com/docs/v5.3/core/timeseries-collections>

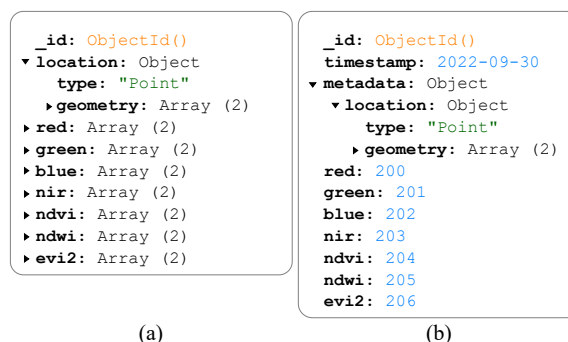


Figura 3. Idealizações para uma modelagem (a) tradicional e (b) *Time Series*.

#### 4. Conclusões

De uma forma geral, a modelagem da Figura 3a atende melhor consultas que desejam recuperar toda a série para um determinado *pixel*. Já a modelagem da Figura 3b atende melhor consultas que desejam recuperar o valor do *pixel* em determinada data.

A próxima etapa desse trabalho será realizar testes de desempenho, tanto de carga, quanto de consulta utilizando as duas modelagens citadas. Após os testes, haverá subsídios suficientes para decidir entre uma e outra. E então, será desenvolvida uma aplicação web que permitirá o acesso a tais dados, de um forma simples e amigável, sem a necessidade de codificação.

Hoje, tanto o BDC quanto outras plataformas em nuvem, como o *Google Earth Engine*, fornecem os dados, mas há uma dificuldade em acessá-los. Adicionalmente, a aplicação desenvolvida será direcionada para o setor da cafeicultura, com o propósito de apresentar os dados de maneira mais acessível para os pesquisadores, extensionistas e, sobretudo, os produtores.

Espera-se que o desenvolvimento desse trabalho contribua para diversos estudos que envolvem a cafeicultura. Por exemplo, o NDWI possibilita o monitoramento dos recursos hídricos e pode contribuir para um controle da irrigação de maneira sustentável. Alterações no NDVI e o EVI2 podem sugerir surgimento de pragas e doenças, apontando para o produtor o local mais afetado e, assim, permitindo diminuir o uso de defensivos agrícolas. Acredita-se que a agricultura pode avançar em direção à sustentabilidade ao fazer uso de dados de sensoriamento remoto. Nesse contexto, a iniciativa proposta neste trabalho busca impulsionar esse avanço na cafeicultura de Minas Gerais.

#### Referências

Souza, V. C. O., Parede, D. A., Volpato, M. M. L., and Alves, H. M. R. (2019). Aplicações do google earth engine na cafeicultura do sul de minas gerais. In *X Simpósio de Pesquisa dos Cafés do Brasil*.

Vieira, T. G. C., Alves, H. M. R., LACERDA, M., VOLPATO, M., and de SOUZA, V. (2010). Estudo espaço-temporal da cafeicultura na região de são sebastião do paraíso-mg, utilizando geotecnologias. In *Reunião Brasileira de Conservação do Solo e da Água*, volume 18.