

Integration of Google Earth Engine and Google Colaboratory applied to digital processing of remote sensing images

Sabrina de Oliveira Gomes , Ana Lúcia B. Candeias

Mestrando em Ciências Geodésicas e Tecnologias da Geoinformação pela Universidade Federal de Pernambuco, Avenida da Arquitetura, s/n - Cidade Universitária, Recife - PE, Brasil. Email: o.sabrina.gomes@gmail.com
Professora do Departamento de Engenharia Cartográfica e Agrimensura, Universidade Federal de Pernambuco.
Received 22 July; accepted 17 September.

ABSTRACT

The application of remote sensing techniques enjoys a variety of fields of study thanks to the wide range of information that can be extracted from the collected data. Obtaining this information is directly influenced by the approaches and methodologies used in processing the acquired data. Currently, sensors provide a substantial volume of data, which can be used in various research contexts and practical applications. Therefore, the present work is driven by the opportunity to develop automated procedures that simplify the processing of images captured by remote sensors, with an application to data from the Landsat-8 satellite. For this, computational routines were developed in Python language, using the Google Earth Engine and Google Colaboratory platforms. These routines cover the application of the Normalized Difference Vegetation Index (NDVI), Soil-Adjusted Vegetation Index (SAVI), Normalized Difference Water Index (NDWI), Normalized Difference Built Area Index (NDBI). The process encompasses a sequence of steps, which begins with the delimitation of the area of interest, followed by the preparation of the images, including the application of masks to remove unwanted elements, such as clouds. And finally, the conversion of imaged information into tabular formats, with statistical, graphical analyzes and cartographic representation of the results complementing the analysis. In summary, this work aims to optimize the analysis of remote sensing images, making the processing stage more efficient and aiding the interpretation of spatial data.

Keywords: Cartography; Automated processes; Python; Landsat-8; Google Earth Engine; Google Collaboration.

Integração do Google Earth Engine e Google Colaboratory aplicados ao processamento digital de imagens de sensoriamento remoto

RESUMO

Os usuários de imagens de sensoriamento remoto têm disponibilidade de acesso a vários bancos de dados de forma gratuita. Esses bancos de imagem com suas respectivas resoluções espacial, espectral, radiométrica e temporal são formados por grande número de dados em nuvem de computadores. As utilizações são voltadas para diversos contextos de pesquisa e aplicações práticas relacionadas ao meio ambiente. Trabalhar com esses dados sem precisar baixá-los localmente se apresenta como vantagem cada vez mais necessária, com redução de tempo de processamento e espaço nos computadores locais. O presente trabalho exemplifica procedimentos com uso desses bancos e o processamento dessas imagens usando imagens do OLI do Landsat-8 e rotinas computacionais em linguagem Python, empregando as plataformas do Google Earth Engine e Google Colaboratory. Essas rotinas abrangem a aplicação do Índice de Vegetação por Diferença Normalizada (NDVI), Índice de Vegetação Ajustado ao Solo (SAVI), Índice de Água de Diferença Normalizada (NDWI), Índice de área construídas por diferença normalizada (NDBI). O processo engloba uma sequência de etapas, que se inicia com a delimitação da área de interesse, seguida pela preparação das imagens, incluindo a aplicação de máscaras para remover elementos indesejados, como nuvens. E por fim, a conversão das informações imageadas em formatos tabulares, com análises estatísticas, gráficas e a representação cartográfica dos resultados complementando a análise. Em resumo, esse trabalho visa otimizar a análise de imagens de sensoriamento remoto em nuvem, tornando mais eficaz a etapa do processamento e auxiliando a interpretação de dados espaciais.

Palavras-Chave: Cartografia; Processos automatizados; Python; Landsat-8; Google Earth Engine; Google Colab.

1. Introdução

Muitas análises de problemas ambientais utilizam dados de Sensoriamento Remoto, Processamento de Imagens e Geoprocessamento.

E necessitam de múltiplas imagens, múltiplos sensores, múltiplas épocas entre outros nessa investigação. Usar ferramentas que trabalhem em

nuvem com bancos de imagens e que ofereça programação em código aberto, é um caminho para a geração de resultados.

Existem algumas opções gratuitas onde o usuário desenvolve seus códigos e compartilha com a comunidade acadêmica e profissional, como é o exemplo do Google Earth Engine (GEE) (ALMEIDA, 2023). Esse artigo de uma forma geral exemplifica o desenvolvimento desses códigos.

O GEE e o Google Colaboratory (Google Colab) são ferramentas que podem auxiliar no desenvolvimento de programação em processamento digital de imagens. O GEE é baseado em uma plataforma de processamento de dados geoespaciais em nuvem. Com isso pode-se ter o acesso e o processamento de uma grande variedade de dados de sensoriamento remoto de diferentes fontes. Já o Google Colab é uma plataforma de desenvolvimento em nuvem baseada em notebooks do Jupyter (DA SILVA, 2020), e possibilita a criação de scripts em Python (DA SILVA, 2020), uso de bibliotecas de processamento de imagens, e dados geoespaciais. A integração dessas duas ferramentas, GEE e Google Colab, podem auxiliar uma análise mais eficiente em grandes volumes de imagens, possibilitando a criação de rotinas.

Nesse artigo serão apresentadas as funcionalidades das plataformas do GEE e do Google Colab, com o objetivo de exemplificar a análise de imagens de sensoriamento remoto, utilizando o Landsat-8, tornando o processo mais otimizado. Para exemplificar, utilizou-se o GEE e Colab para se obter os índices NDVI, SAVI, NDWI e NDBI e foram aplicados ao município de Petrolina – PE. A seguir tem-se a definição dos índices que foram usados.

Índice de vegetação por diferença normalizada (NDVI)

O NDVI (Normalized Difference Vegetation Index) é um índice de sensoriamento remoto amplamente utilizado (Cohen et al., 2003) para avaliar e quantificar a presença de vegetação, em uma determinada área. Esse índice é útil em aplicações de monitoramento ambiental, estudos agrícolas, gestão de recursos naturais e análises climáticas (Equação 1):

$$NDVI = (NIR - RED) / (NIR + RED) \quad (1)$$

Onde:

NIR: reflectância na banda do infravermelho próximo;

RED: reflectância na banda do vermelho;

Os valores de NDVI variam de -1 a +1, com valores positivos indicando maior presença de vegetação saudável e valores negativos indicando áreas sem vegetação ou com pouca vegetação.

Índice de vegetação ajustado ao solo (SAVI)

O SAVI (Índice de vegetação ajustado ao solo), desenvolvido por Huete (1988), e tem como objetivo melhorar os valores encontrados no NDVI, a equação se diferencia do NDVI no termo $(1 + L)$, onde temos L como uma constante, é possível calcular conforme a equação (2):

$$SAVI = (1 + L) \times (NIR - R) / (NIR + R + L) \quad (2)$$

Onde:

NIR: é a banda que corresponde a reflectância do infravermelho próximo;

R: é a banda que corresponde a reflectância do vermelho;

L: é um fator de ajuste.

O fator L pode apresentar valores variáveis, segundo Huete (1988), os valores podem variar de 0 a 1, onde para densidades baixas de vegetação podem utilizar o valor 1, para médias densidades de vegetação o valor 0,5 e para altas densidades de vegetação o valor 0,25.

O fator L é considerada uma constante empírica que possui a finalidade de minimizar a sensibilidade do índice de vegetação as variações de refletância dos tipos do solo.

Índice de água de diferença normalizada (NDWI)

O NDWI (Índice de Água de Diferença Normalizada) é utilizado para destacar as características da água, dessa forma, a reflectância do solo e da vegetação é consideravelmente reduzida, permitindo que a massa de água se destaque (MCFEETERS, 1996).

Segundo Jackson et. al. (2004) afirmam que, o NDVI saturava ao verificar mudanças na biomassa e nos estresses de umidade das plantas, enquanto, a representação do NDWI apresentava melhor as mudanças no conteúdo de umidade das vegetações.

O NDWI é calculado pela diferença normalizada entre o infravermelho próximo (NIR) e o verde (Green) dividido pela soma do infravermelho próximo (NIR) e o verde (Green), é possível calcular segundo a equação (3):

$$NDWI = (GREEN - NIR) / (GREEN + NIR) \quad (3)$$

Onde:

NIR: reflectância na banda do infravermelho próximo;

GREEN: reflectância na banda do verde;

Seus resultados variam de -1 a 1, onde os valores positivos indicam a presença de água, enquanto valores negativos indicam a presença de vegetação, McFeeters (1996) definiu zero como o limiar.

Índice de área construída por diferença normalizada (NDBI)

O NDBI (Índice de área construída por diferença normalizada), é um índice utilizado para destacar o adensamento de construções e machas urbanas (Zha et al., 2003), é possível calcular conforme a equação (4):

$$NDBI = (IVM - IVP) / (IVM + IVP) \quad (4)$$

Onde:

IVM: é a banda que corresponde ao infravermelho de ondas curtas ou médio;

IVP: é a banda que corresponde ao infravermelho próximo;

Esses índices são úteis para monitorar a saúde da vegetação, monitoramento da água e do crescimento urbano, contribuindo para diversas aplicações relacionadas ao meio ambiente e ao uso sustentável dos recursos naturais.

2. Material e métodos

Para exemplificar o uso do GEE e o Google Colab e as equações relativas aos Índice de Vegetação por Diferença Normalizada (NDVI), Índice de Vegetação Ajustado ao Solo (SAVI), Índice de Água de Diferença Normalizada

(NDWI), Índice de área construídas por diferença normalizada (NDBI). A Figura 2 apresenta o diagrama de blocos sobre as etapas desenvolvidas.

Caracterização da área de estudo

Utilizou-se o município de Petrolândia que está localizada no estado de Pernambuco, na região Nordeste do Brasil (Figura 1) para exemplificação da metodologia apresentada na Figura 2.

Petrolina é município do sertão de Pernambuco, e que possui uma área territorial de aproximadamente 1056,6 km², localizado as margens do Rio São Francisco, à uma distância aproximada de 430km da capital Recife (IBGE, 2022). De acordo com o IBGE (2022), sua população é aproximadamente de 36.548 habitantes. Este município está situado a 319 metros de altitude, limita-se ao norte com município de Floresta, ao sul com município de Jatobá e o estado da Bahia, a oeste com o estado da Bahia e a Leste com município de Tacaratu, e possui as coordenadas geográficas: Latitude 9° 4' 12'' Sul, Longitude 38° 16' 54'' Oeste.

O clima é predominantemente quente. Caracteriza-se como região semiárida, com uma temperatura média anual em torno de 25°C, a estação chuvosa concentrada no período de janeiro, fevereiro e março, enquanto os períodos de seca se concentram em setembro, outubro e novembro. A vegetação é predominantemente a caatinga, que se caracteriza por possuir espécies vegetais com alta capacidade de retenção de água. Essas condições climáticas influenciam diretamente na vegetação predominante na região e nas atividades econômicas locais.

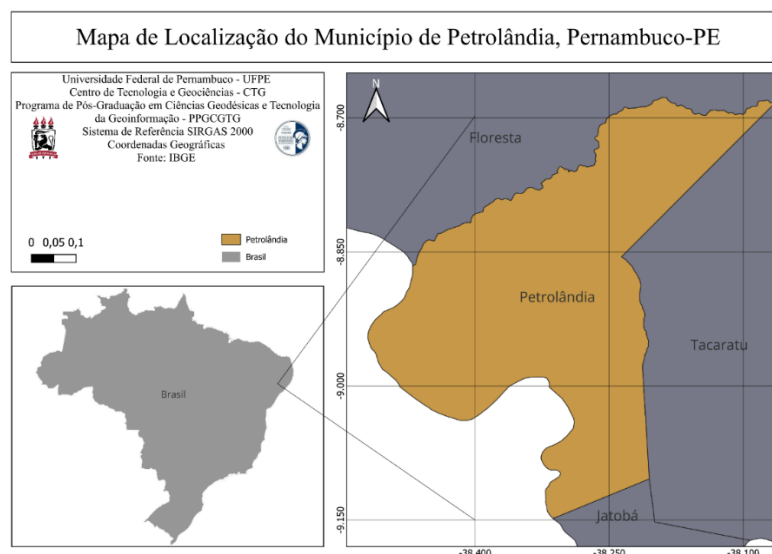


Figura 1 – Mapa Localização da área de estudo, cidade do município de Petrolândia - PE, Brasil

2.1 Materiais e Procedimento Metodológicos

Os materiais aqui apresentados serão divididos em dois grupos, esses grupos foram intitulados de dados e tecnologias.

Materiais considerados como dados:

- I. Imagens de satélite do Landsat-8 na faixa do visível e infravermelho próximo em formato .tiff;
- II. Shapefile do IBGE, com a delimitação da área de interesse.

Materiais considerados como tecnologias:

- I. Desenvolvimento do trabalho em linguagem Python e utilização das bibliotecas;
- II. Utilização do GEE e Google Colaboratory.

O procedimento metodológico é descrito a partir do fluxograma apresentado na Figura 2.

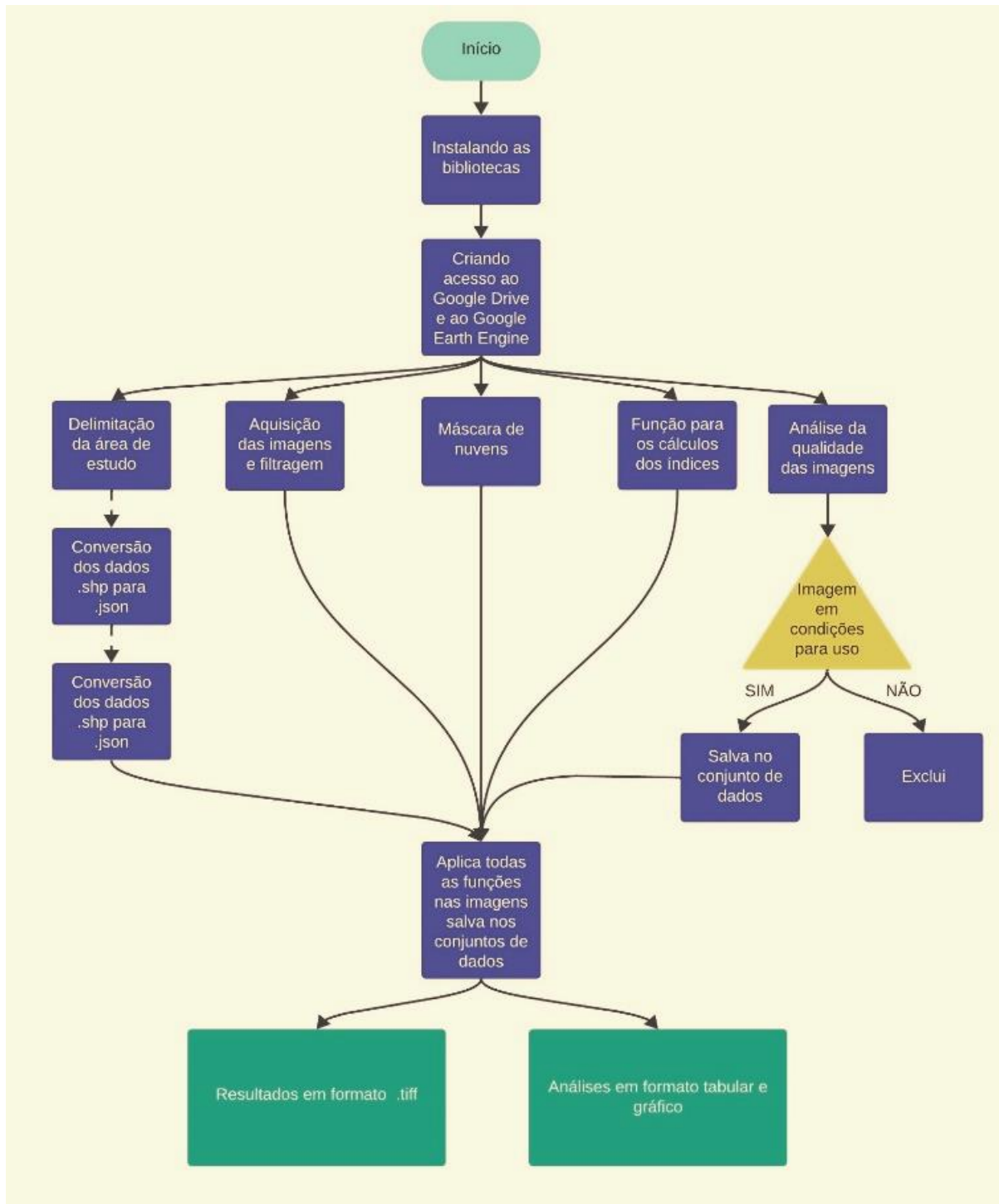


Figura 2 – Fluxograma da metodologia aplicada.

2.1.1 Desenvolvimento

A primeira etapa desempenha um papel essencial ao preparar o ambiente de execução do código, onde é feita toda a configuração do acesso e instalação das bibliotecas necessárias

Essa fase é responsável por instalar e importar as bibliotecas que permitem a interação com os dados e todas as funcionalidades do GEE, além de possibilitar recursos de visualização e manipulação das informações geoespaciais em formato DataFrame.

Com as bibliotecas devidamente instaladas, é necessário estabelecer o acesso ao Google Drive, é crucial que essa etapa seja respeitada, pois ela permite que seja integrada o ambiente de execução do Colab ao Drive, onde o usuário realiza a autenticação e acessa todos os arquivos armazenados, fazendo toda a atividade diretamente no código, facilitando o acesso direto e o armazenamento de qualquer resultado gerado pelo código.

Ambas as etapas são essenciais para garantir a preparação adequada do ambiente de execução, proporcionando todas as ferramentas necessárias, então iniciar de forma estruturada e integrada contribui para eficácia e conveniência do processo.

A delimitação da área de estudo realizada dentro do ambiente do Colab necessita que seja carregado um arquivo que contenha os dados geoespaciais (shapefile - .shp) que estão localizados no Google Drive, após isso, as informações do shapefile precisam ser convertidas em formato .json. Ambos são arquivos vetoriais geoespaciais, a diferença é que o .json permite ser aplicado em plataformas web. Portanto, é exibido esses dados em um mapa interativo usando a biblioteca geemap no ambiente Colab, a exibição permite que o usuário identifique se a delimitação ocorreu como esperado.

A base de dados do GEE abriga imagens desde a operação do Landsat-1 até o Landsat-9. Os dados gerados pelo Landsat são organizados em coleções, que determinam o nível de qualidade da imagem, tendo uma consistência garantida no uso destas, independente dos instrumentos utilizados na aquisição (Landsat Missions, [S.D.]).

Foram utilizadas imagens fornecidas pelo Landsat-8. Tais imagens são pertencentes a coleção 2, nível 1. Com a definição do conjunto a ser utilizado, o próximo passo consiste em filtrá-lo por área e período, Figura 3.

```
#Importando imagem do Landsat 8 collection e aplicando filtros

l8 = ee.ImageCollection('LANDSAT/LC08/C02/T1_L2')\
    .filterBounds(region)\
    .filterDate('2023-01-01','2023-12-31')\
    .map(maskL8sr)\
    .map()
```

Figura 3 – Leitura dos dados.

O landsat-8 é um satélite que captura imagens em várias bandas espectrais, incluindo bandas ópticas (visíveis e infravermelhas) e bandas térmicas. Entretanto, em algumas imagens, podem ocorrer a presença de nuvens, que podem prejudicar a interpretação dos objetos terrestres, especialmente quando se deseja analisar alterações ou caracterizar uma área.

A etapa consistiu em aplicar um código intitulado de máscara de nuvens nas imagens do landsat-8 de forma a preparar as imagens que serão utilizadas. O propósito é eliminar os pixels afetados por nuvens e sombras, permitindo uma melhor interpretação e análise das informações contidas na imagem a superfície da Terra.

```
##Máscara de nuvens

def maskL8sr(image):

    qaMask = image.select('QA_PIXEL').bitwise_and(int('11111',
2)).eq(0)

    saturationMask = image.select('QA_RADSAT').eq(0)

    opticalBands =
image.select('SR_B.').multiply(0.0000275).add(-0.2)

    thermalBands =
image.select('ST_B.*').multiply(0.00341802).add(149.0)

    return image.addBands(opticalBands, None, True)\
        .addBands(thermalBands, None, True).updateMask(qaMask)\
        .updateMask(saturationMask)
```

Figura 4 – Máscara de nuvens.

Para realizar o cálculo NDVI, SAVI, NDWI e NDBI foram criadas funções que recebem uma imagem de satélite, e assim realiza-se o cálculo para essa imagem, adiciona as bandas necessárias, recorta a imagem baseada na região específica e mantém as propriedades importantes das bandas na imagem resultante.

Para que o código funcione de forma correta, é importante que a imagem que vai ser utilizada como entrada tenham as bandas SR_B5 e SR_B4 para o cálculo do NDVI, enquanto que para o SAVI foram necessárias as bandas SR_B5 e SR_B4, enquanto para o NDWI é necessário que tenha as seguintes bandas SR_B5 e SR_B3, para o NDBI é necessário que tenha as seguintes bandas SR_B5 e SR_B6. A variável da região deve estar corretamente definida como a área do recorte.

```
##Função NDVI (ÍNDICE DE VEGETAÇÃO NORMALIZADA)

def NDVI (img):

    Ndvi_image = img.normalizedDifference(['SR_B5',
    'SR_B4']).rename('NDVI')

    return
img.addBands(Ndvi_image).clip(region).copyProperties(img,
["system:time_start"])
```

Figura 4 – Função para calcular NDVI.

```
# Função SAVI

def SAVI(img):

    savi = img.expression(
        '(1 + L) * (NIR - RED) / (NIR + RED + L)',
        {
            'NIR': img.select('SR_B5'), # Banda 5:
            Infravermelho próximo
            'RED': img.select('SR_B4'), # Banda 4:
            Vermelho
            'L': 0.5 # Fator de ajuste
        }
    ).rename('SAVI')

    # Adicionando a banda SAVI
    img = img.addBands(savi)

    # Recortando na região de interesse
    img = img.clip(region)

    # Copiando propriedades e adicionando a data formatada
    date =
    ee.Date(img.get('system:time_start')).format('YYYY-MM-dd')
    img = img.set('date', date)

    return img
```

Figura 5 – Função para calcular SAVI.

```
##Função NDWI

def NDWI(img):

    Ndwi_image = img.normalizedDifference(['SR_B3',
    'SR_B5']).rename('NDWI')

    return
img.addBands(Ndwi_image).clip(region).copyProperties(img,
["system:time_start"])
```

Figura 6 – Função para calcular NDWI.

```
def NDBI(img):

    Ndbi_image = img.normalizedDifference(['SR_B5',
    'SR_B6']).rename('NDBI')

    img =
    img.addBands(Ndbi_image).clip(region).copyProperties(img,
    ["system:time_start"])

    date =
    ee.Date(img.get('system:time_start')).format('YYYY-MM-dd')

    img = img.set('date', date)

    return img
```

Figura 7 – Função para calcular NDBI.

Por fim, com os índices calculados, é possível gerar os resultados em formato tabular e gráficos, onde podemos observar os resultados e gerar análises estatísticas.

3. Resultados e discussão

No geral, o código cria uma visualização informativa que permite avaliar os dados da área específica, ajudando a identificar padrões. Abaixo é possível visualizar os resultados obtidos, onde será apresentado os dados com maior e menor índice pluviométrico, com o intuito de apresentar as imagens obtidas em períodos chuvosos e mais secos, de modo a evidenciar os resultados nesses períodos. A Figura 8 exemplifica os resultados para os índices NDVI, SAVI, NDWI e NDBI para duas épocas (24/09/2023 e 26/10/2023).

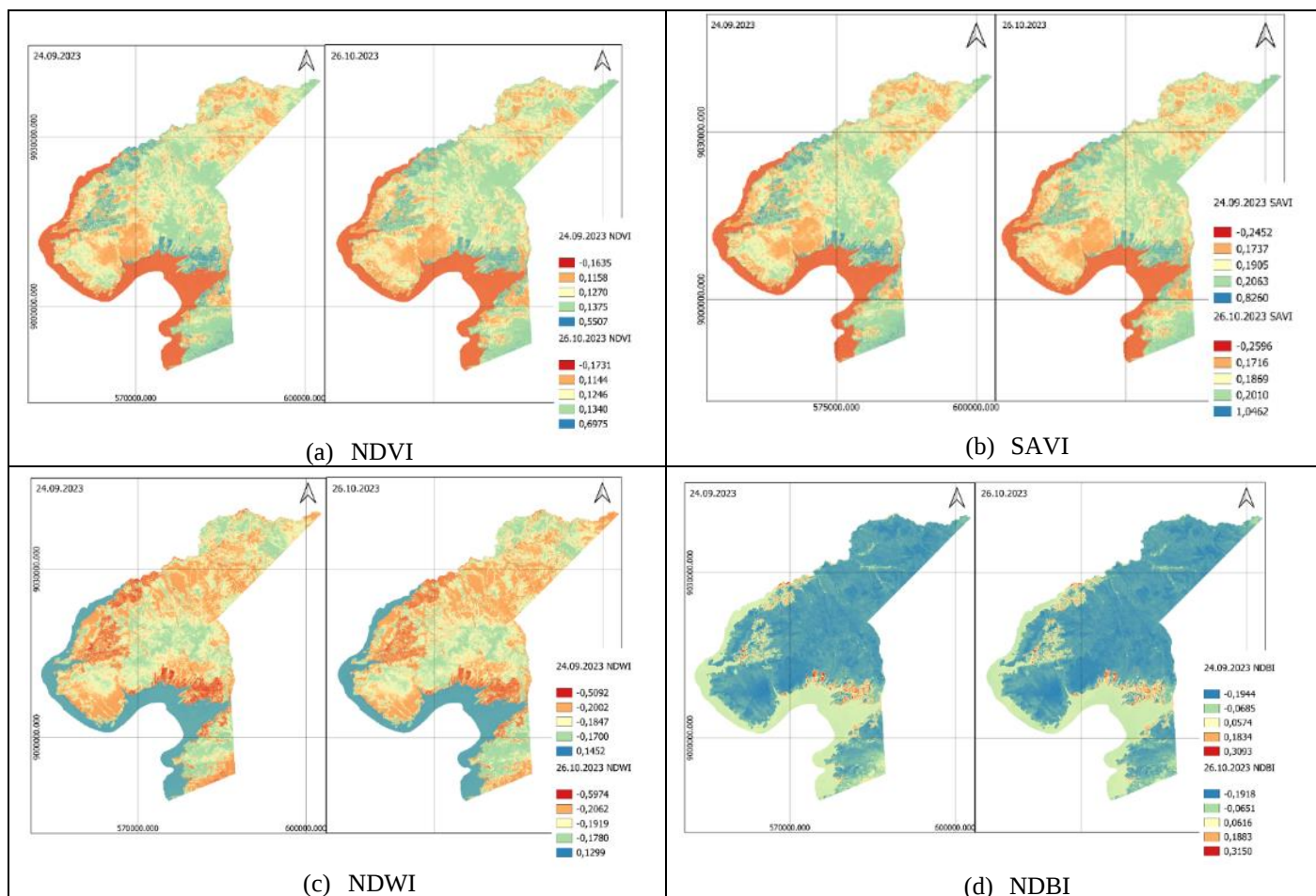


Figura 8 – Resultados para o NDVI, SAVI, NDWI e NDBI.

Conclusões

O trabalho apresentou uma sequência lógica para o processamento e a funcionalidade do código usando GEE e COLAB. Utilizou-se como parâmetros iniciais as definições de:

- área de estudo
- datas de interesse
- sensor

O programa permitiu ao usuário escolher o período de análise, onde o tempo de execução é determinado pelo número de imagens e pixels envolvidos. E também demonstrou eficiência no tempo de execução, beneficiado pelas otimizações do ambiente de execução.

Em resumo podemos observar que esse ambiente de trabalho possibilita o aprimoramento contínuo do trabalho, possibilitando a integração com outros pesquisadores onde podemos compartilhar melhorias e insights valiosos para futuros desenvolvimentos na área.

Os resultados obtidos nesse trabalho mostram-se úteis, principalmente levando em consideração o objetivo da criação de rotinas.

Os métodos aqui apresentados e discutido, formam uma base para a implementação de novos projetos. Uma gama de melhorias e ajustes podem ser feitos buscando um resultado mais satisfatório. O projeto é compartilhável em nuvem e ficará disponível para futuros aprimoramentos.

Referências

- Almeida, K. C. de, Costa, A. P. F. da, & Rocha, T. B. da. (2023). Utilização do Google Earth Engine para análise comparativa da linha de costa de dois deltas da costa leste brasileira entre 1986 e 2021: Delta do rio São Francisco (AL/SE) e delta do rio Paraíba Do Sul (RJ). *Geo UERJ* (42). <https://doi.org/10.12957/geouerj.2023.74650>
- Cohen, W. B., Maersperger, T. K., Gower, S. T., & Turner, D. P. (2003). An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sensing of Environment*, 84, 561-571.
- Huete, A. R. (1988). A Soil Adjusted Vegetation Index (SAVI). *Remote Sensing of Environment*, 25, 295-309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)

- Jackson, T. J., et al. (2004). Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans. *Remote Sensing of Environment*, 92 (4), 475-482.
- McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17 (7), 1425-1432.
- Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24 (3), 583-659.
- Da Silva, M. D. (2020). Aplicação da ferramenta Google Collaboratory para o ensino da linguagem Python. In *Anais da IV Escola Regional de Engenharia de Software* (pp. 67-76). SBC.
- IBGE. (2023). *Área territorial brasileira 2022*. Rio de Janeiro: IBGE.
- IBGE. (2024). *Censo Demográfico 2022, Cadastro Central de Empresas 2022*. Rio de Janeiro: IBGE.