BRAZIL DATA CUBE
brazildatacube.org

Earth observation data cubes and satellite image time series analysis

Karine R. Ferreira

Earth Observation and Geoinformatics Division (DIOTG)

National Institute for Space Research (INPE)

FUNDO AMAZÔNIA

FUNCATE
Fundação de Ciência, Aplicações e Tecnologia Espaciais

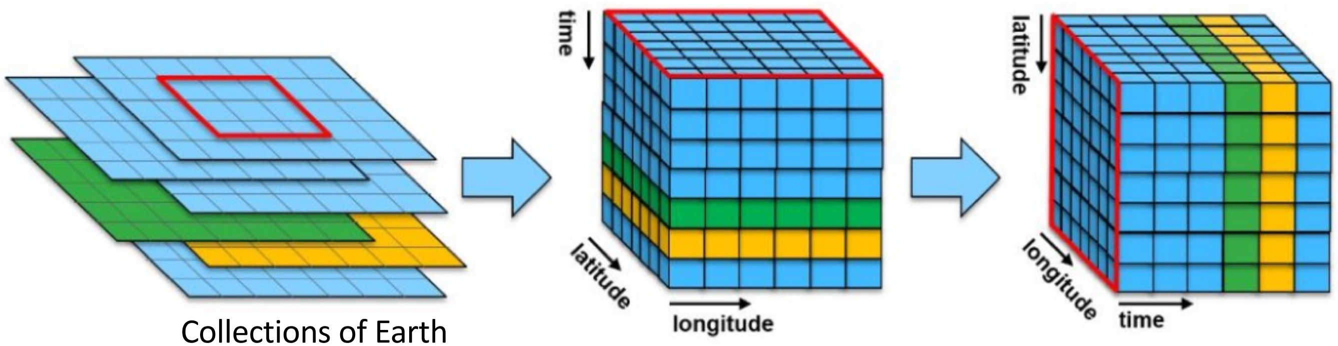
BNDES

INPE

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES

Cubos de Dados de Observação da Terra (Earth observation data cubes)

Cubo de dados de observação da Terra pode ser definido como um ***array* multidimensional de valores** (espaço, tempo, propriedades) usado para descrever uma sequência temporal de imagens (Killough, 2017).



Collections of Earth observation satellite imagens

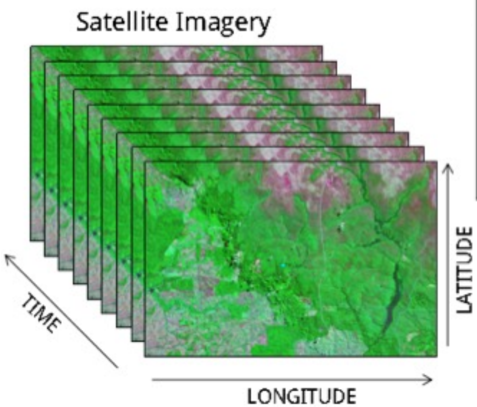
Data cubes –
four-dimensional array.

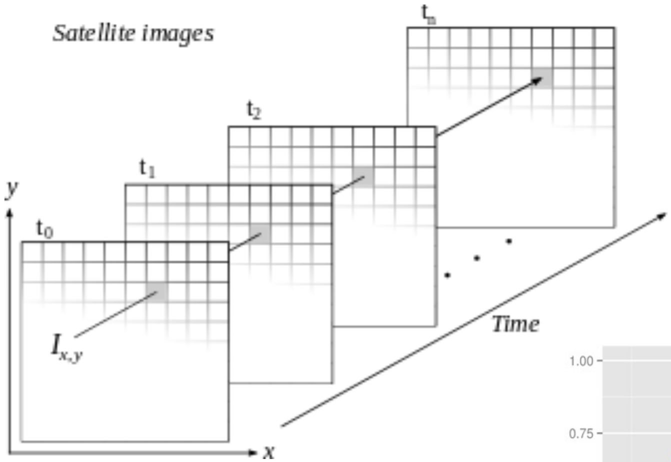
Source: [Kopp et al, 2019]

Termo para referenciar
uma maneira de organizar.

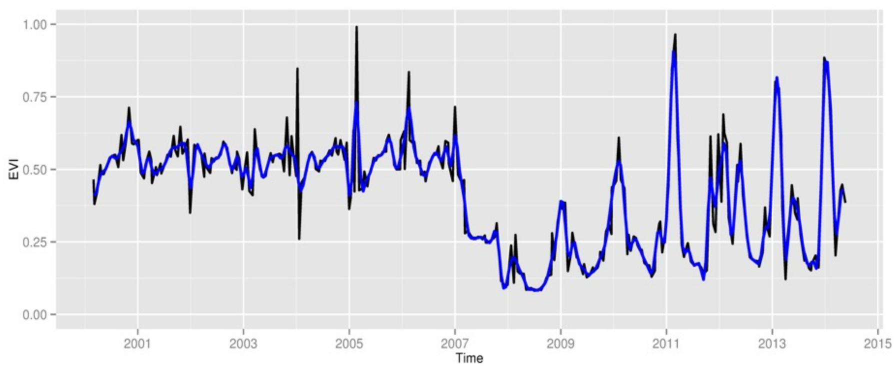
Cubos de Dados de Observação da Terra

Um cubo de dados de observação da Terra pode ser também definido como um conjunto de séries temporais associadas a pixels alinhados espacialmente, prontos para análise. Cubos de dados são criados principalmente para suportar **análise de séries temporais de imagens** (Appel et al., 2019).





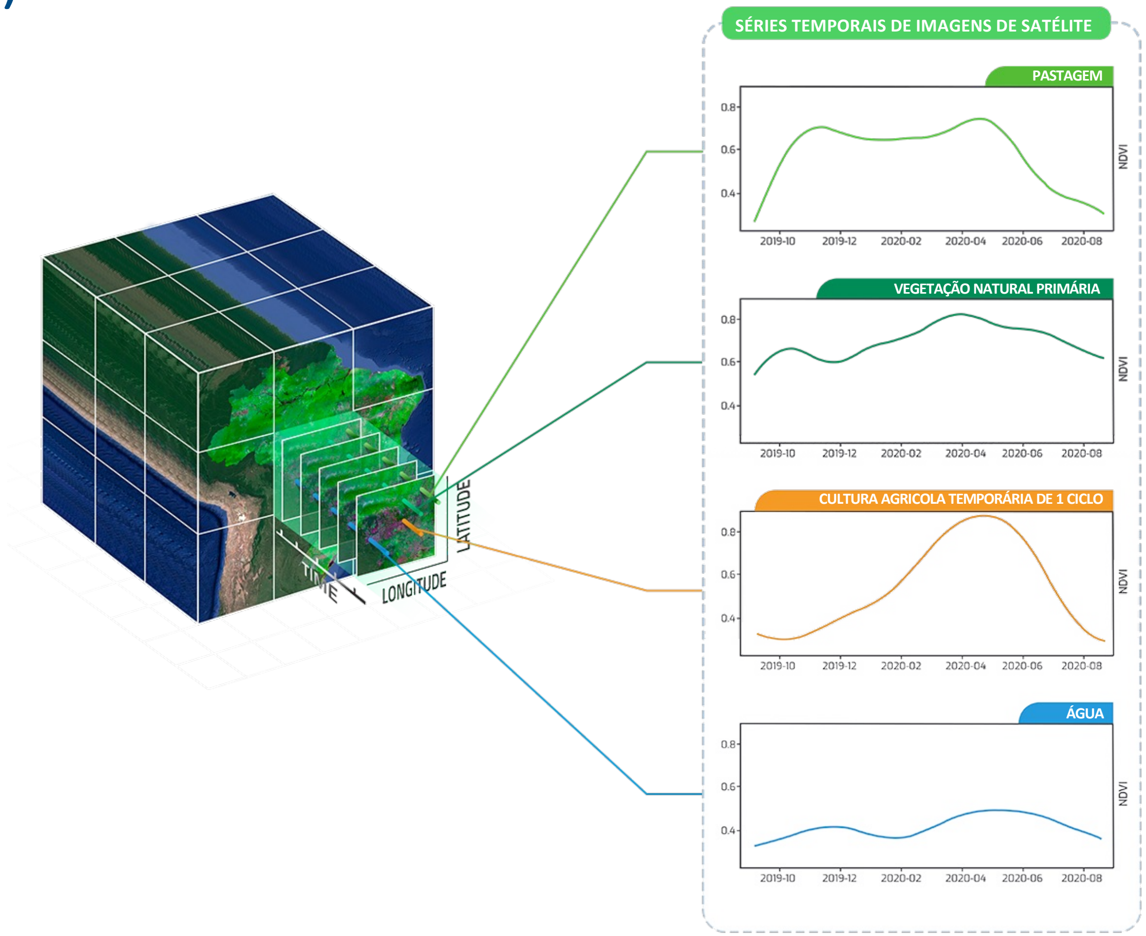
Séries temporais: variação de uma propriedade (espectral ou derivada) ao longo do tempo de uma localização espacial ou pixel.



Why Earth Observation (EO) Data Cubes ?

Multidimensional arrays of satellite images with four dimensions (latitude, longitude, time and attributes), mainly to support **image time series analysis**.

(Appel et al., 2019)



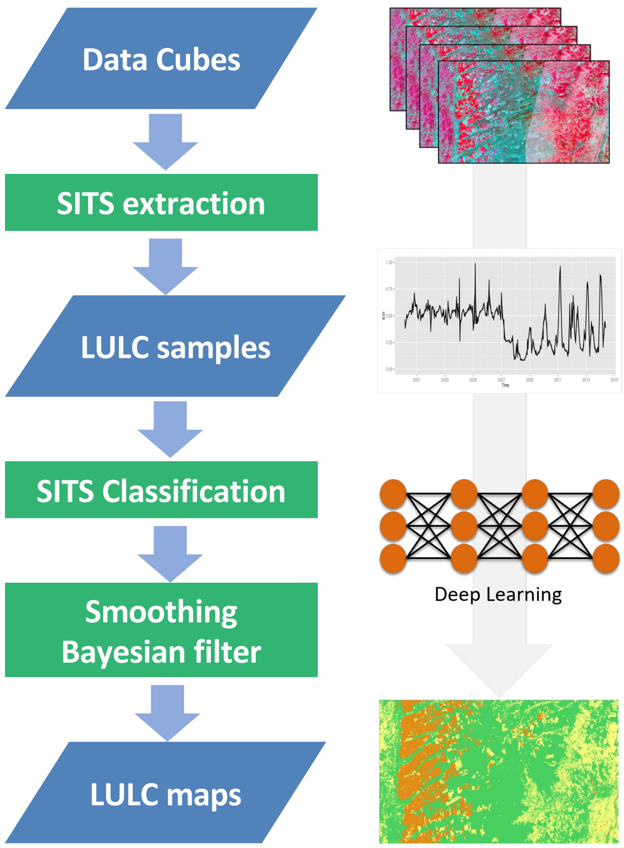
BRAZIL
DATA CUBE

Land use and land cover (LULC) maps

EO data cubes, satellite image time series (SITS) analysis and machine learning to produce LULC maps from big Earth observation data.

SITS reveal complex underlying processes that would be difficult to assess using bi-temporal or even annual change detection approaches.

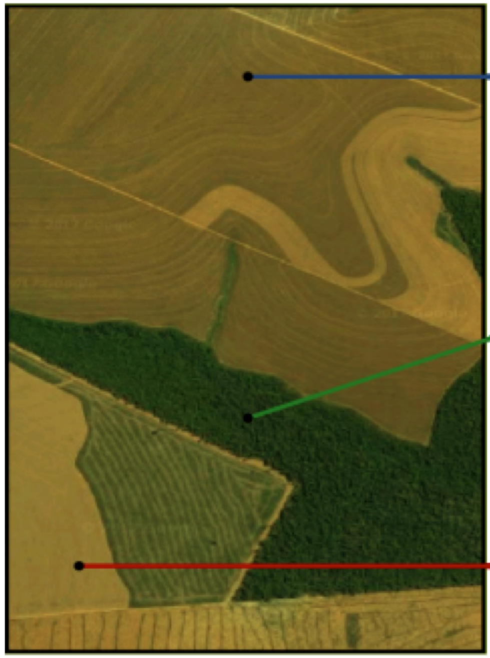
(Pasquarella et al., 2016)



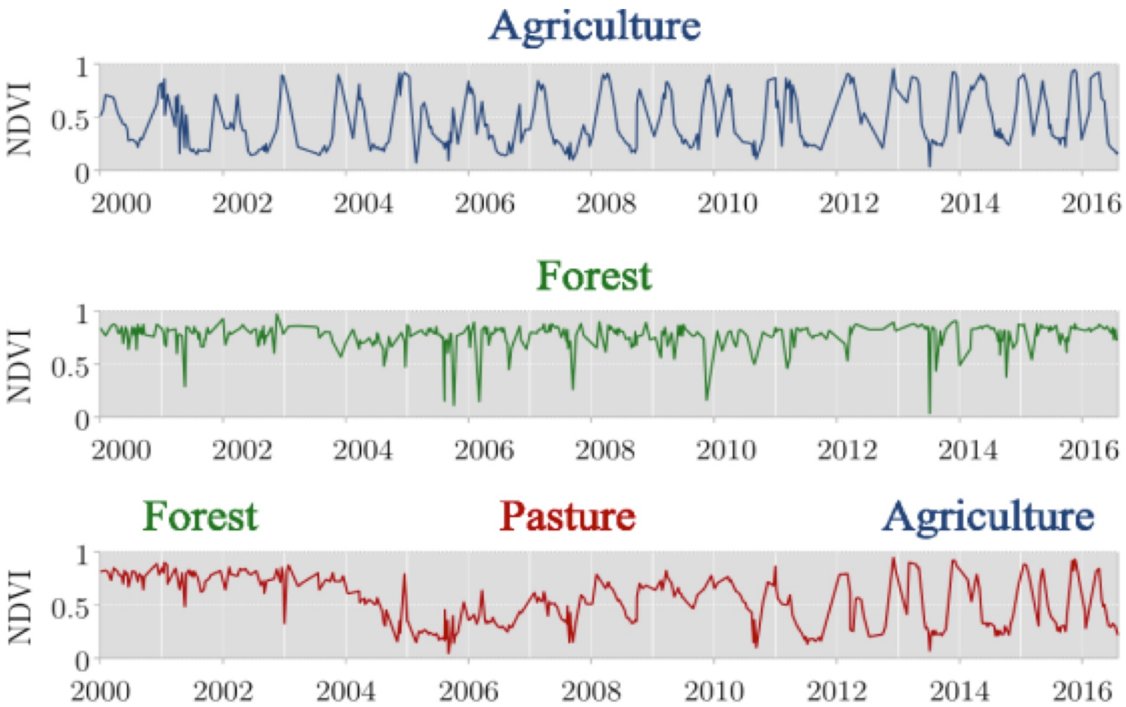
Source: [Ferreira et al, 2020]

Análise de Séries Temporais de Imagens

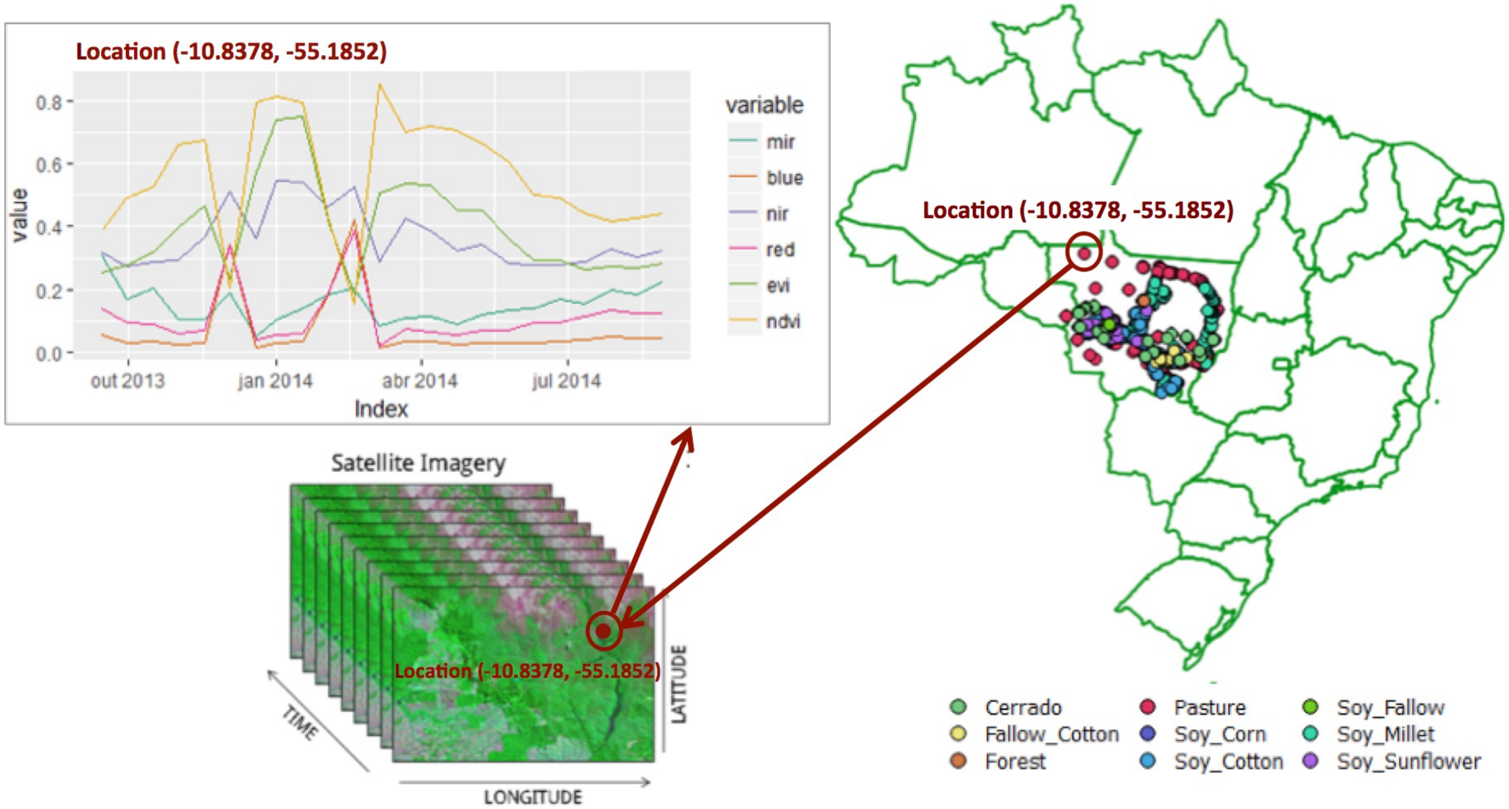
Amostras de uso e cobertura do solo



Três localizações espaciais e suas séries temporais NDVI associadas.

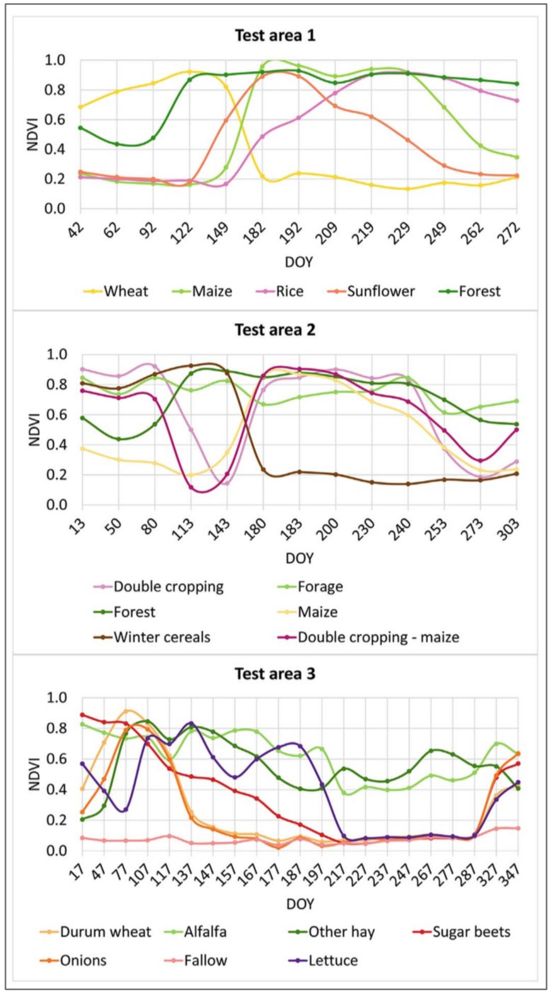


Análise de Séries Temporais de Imagens



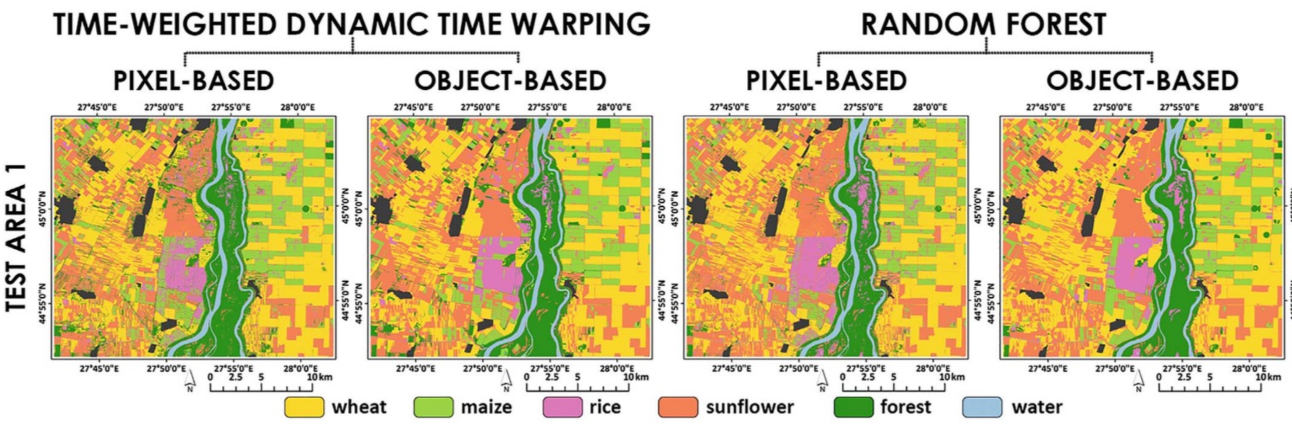
Santos, L.A., Ferreira, K.R., Picoli, M., Camara, G.: Self-organizing maps in earth observation data cubes analysis. **13th International Workshop on Self-Organizing Maps (WSOM).** pp. 70–79 (2019)

Potencial - Análise de séries temporais de imagens



Agricultura

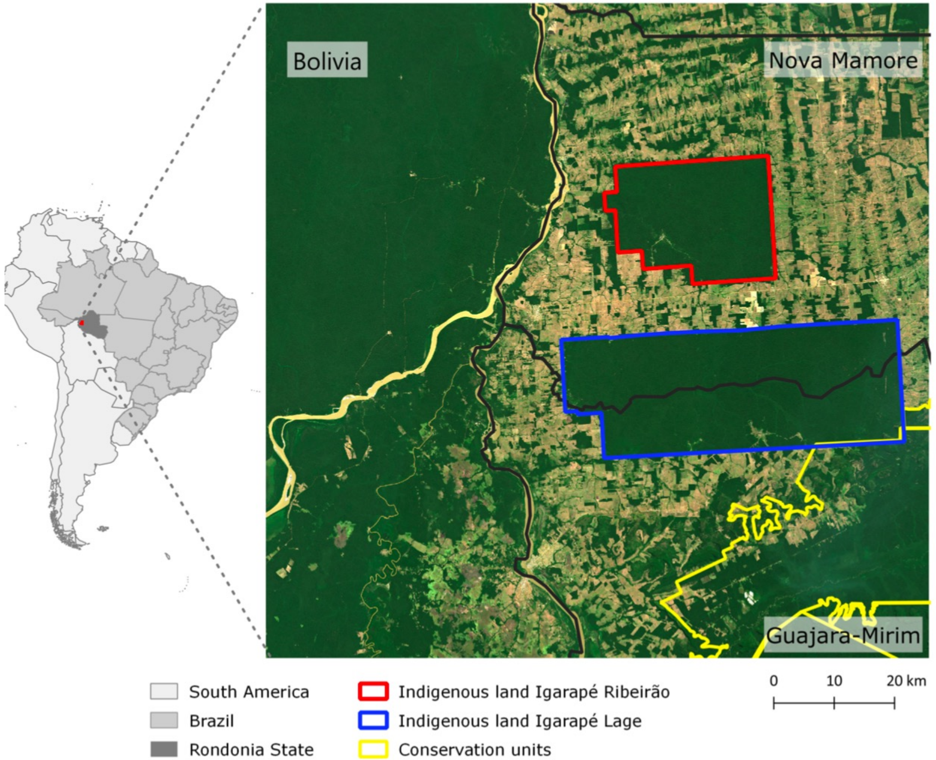
Séries temporais NDVI de imagens Sentinel 2
Tipos de agricultura: Trigo, Milho, Arroz, Girassol, Floresta e Água
Métodos – TWDTW e Ranfon Forest



Source: [Belgiu et al., 2018]

Sentinel-2 cropland mapping using pixel-based and object-based time- weighted dynamic time warping analysis
Remote Sensing of Environment, 2018

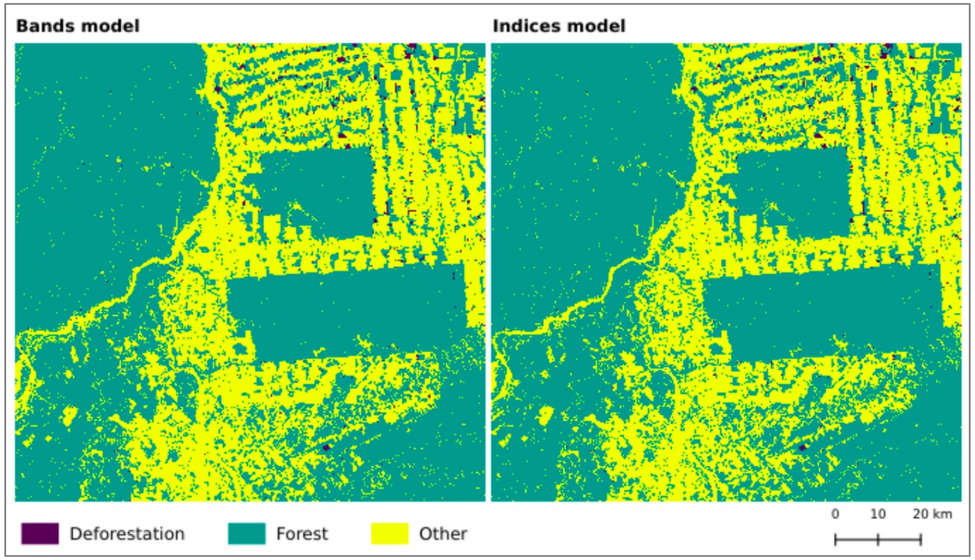
Potencial - Análise de séries temporais de imagens



Source: [Sanchez et al., 2020]

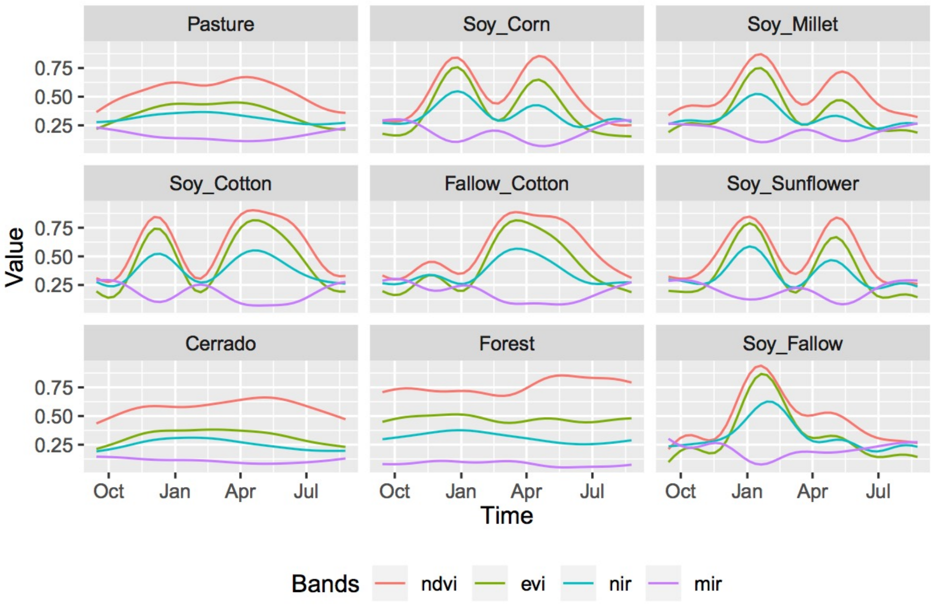
Combining Time Series Analysis with Machine Learning for Detection of Tropical Deforestation with High Accuracy (... Submitted ...)

Desmatamento



Séries temporais de imagens Sentinel 2 (Bandas MSI e Índices NDVI, EVI e NDMI)
Classes: Desmatamento, Floresta e Outros
Métodos – Ranfon Forest

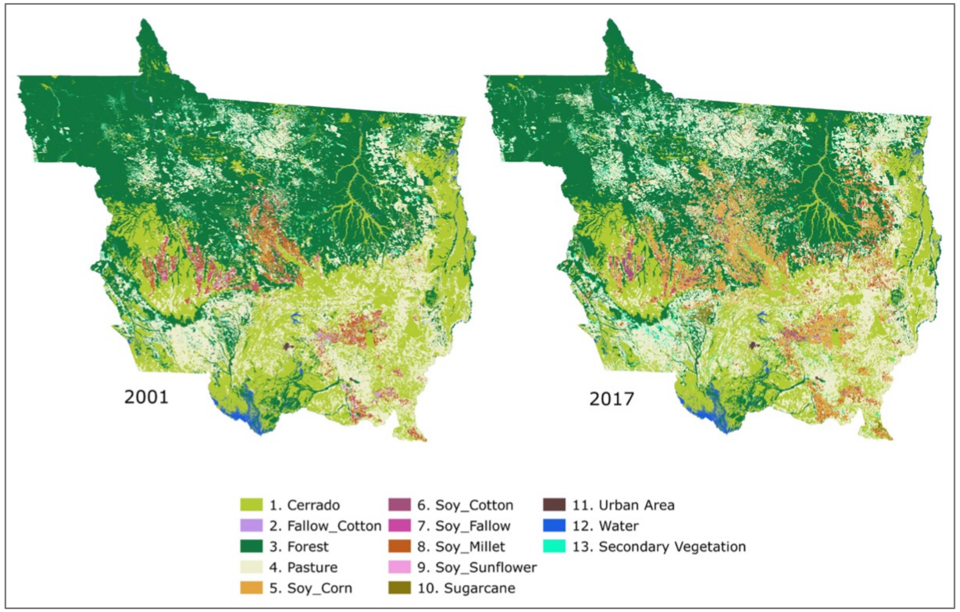
Potencial - Análise de séries temporais de imagens



Source: [Simoes et al., 2020]

Land use and cover maps for Mato Grosso State in Brazil from 2001 to 2017
Scientific Data, 2020

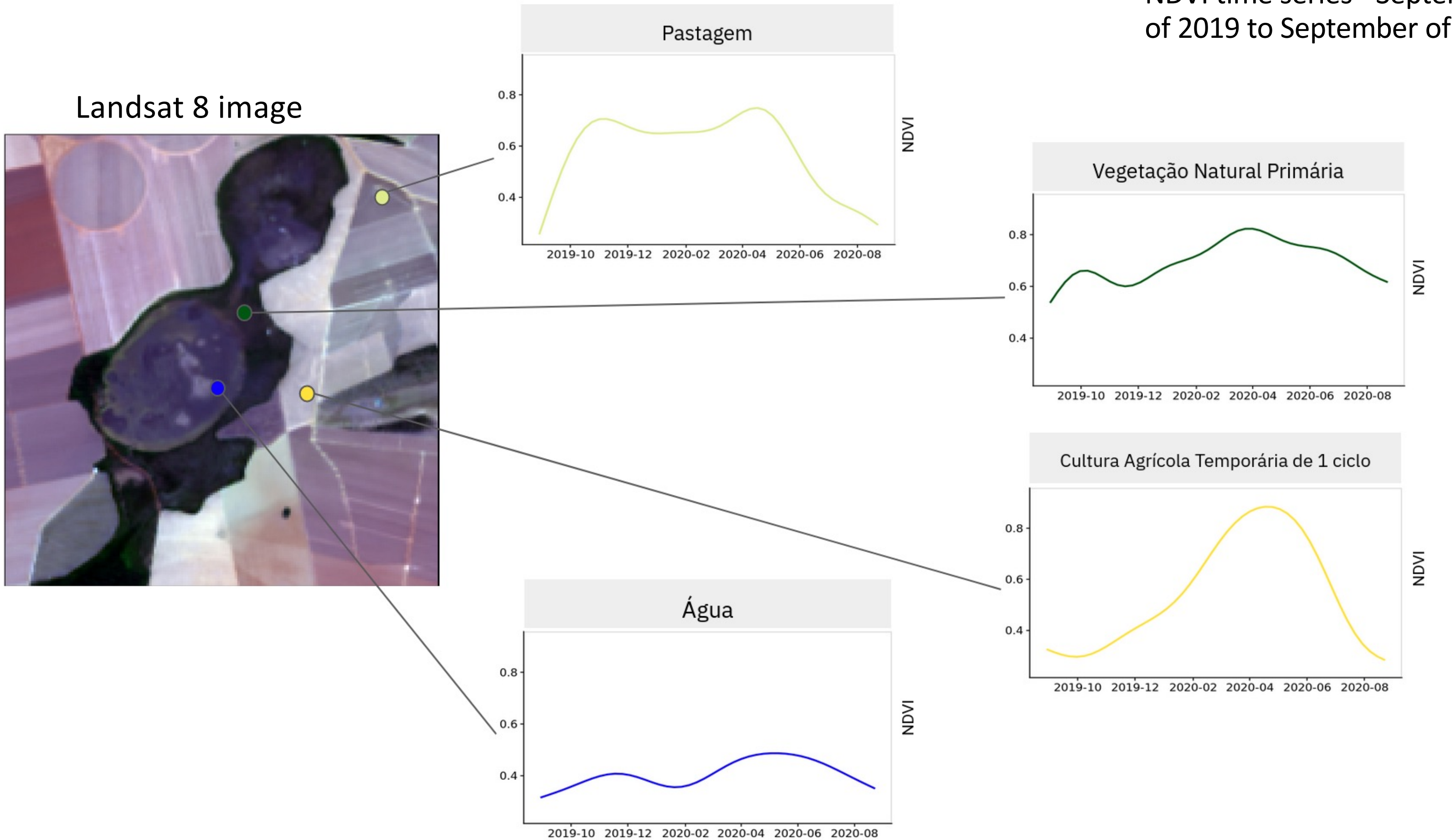
Uso e Cobertura do Solo



Séries temporais NDVI, EVI, NIR, MIR de imagens
MODIS – Produto MOD13Q1
Métodos – SVM (Support Vector Machine)

NDVI time series – Landsat Data Cube (16-days)

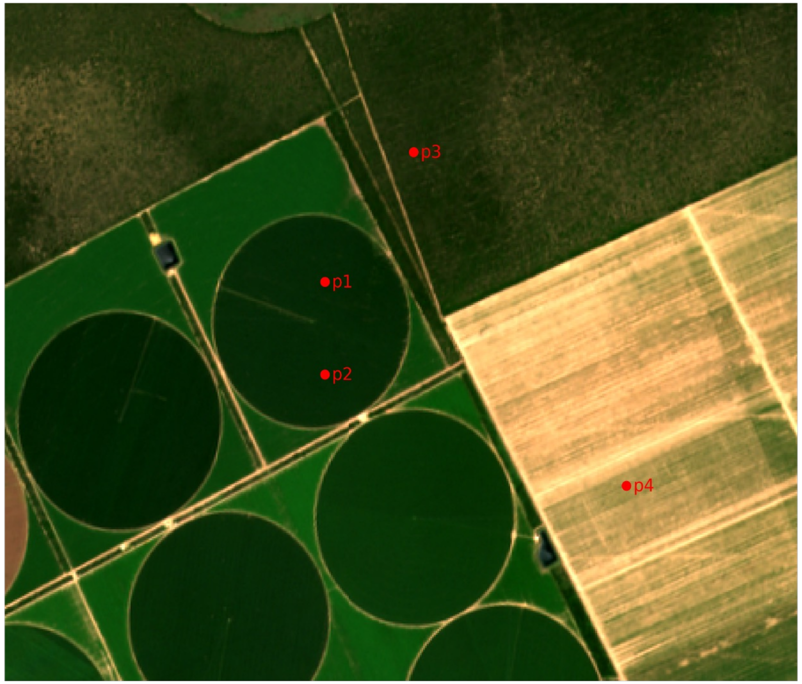
NDVI time series - September of 2019 to September of 2020



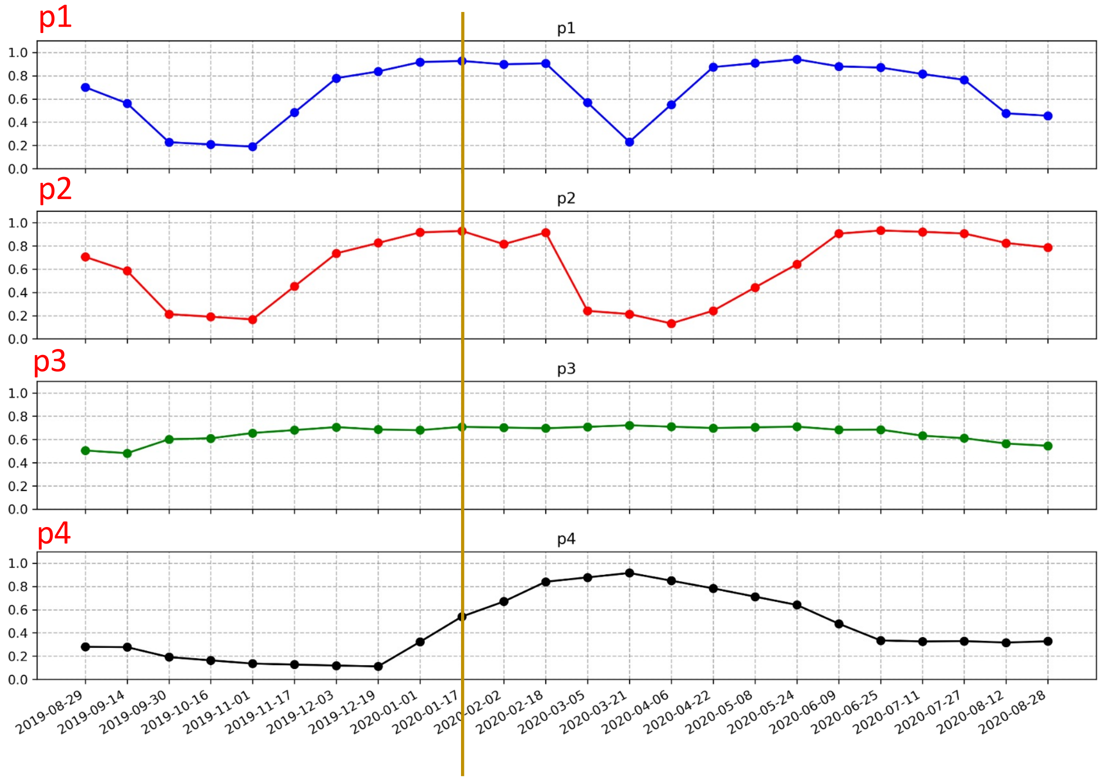
NDVI Time Series – Sentinel-2 Data Cube (16-days)

BRAZIL
DATA CUBE

Sentinel-2 RGB



2020-01-17

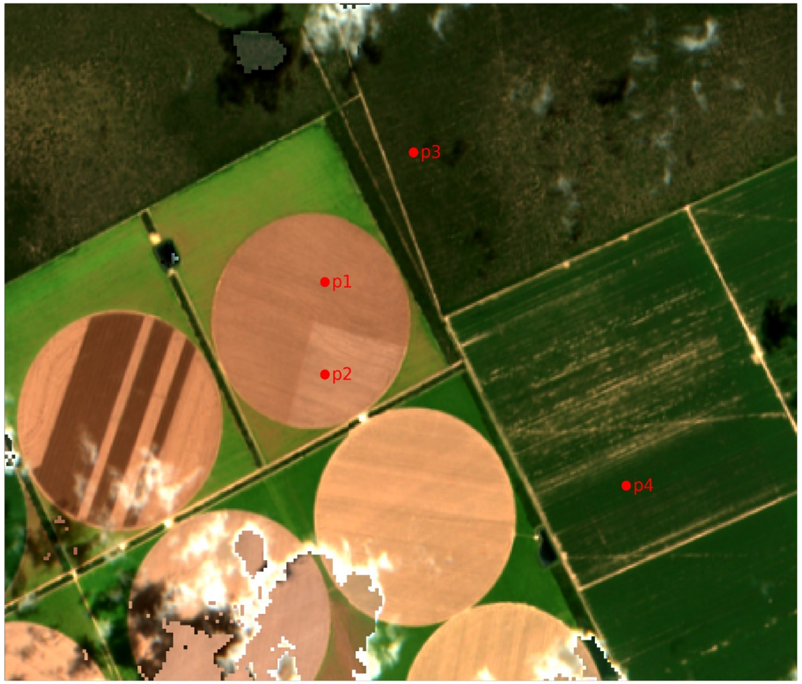


2020-01-17

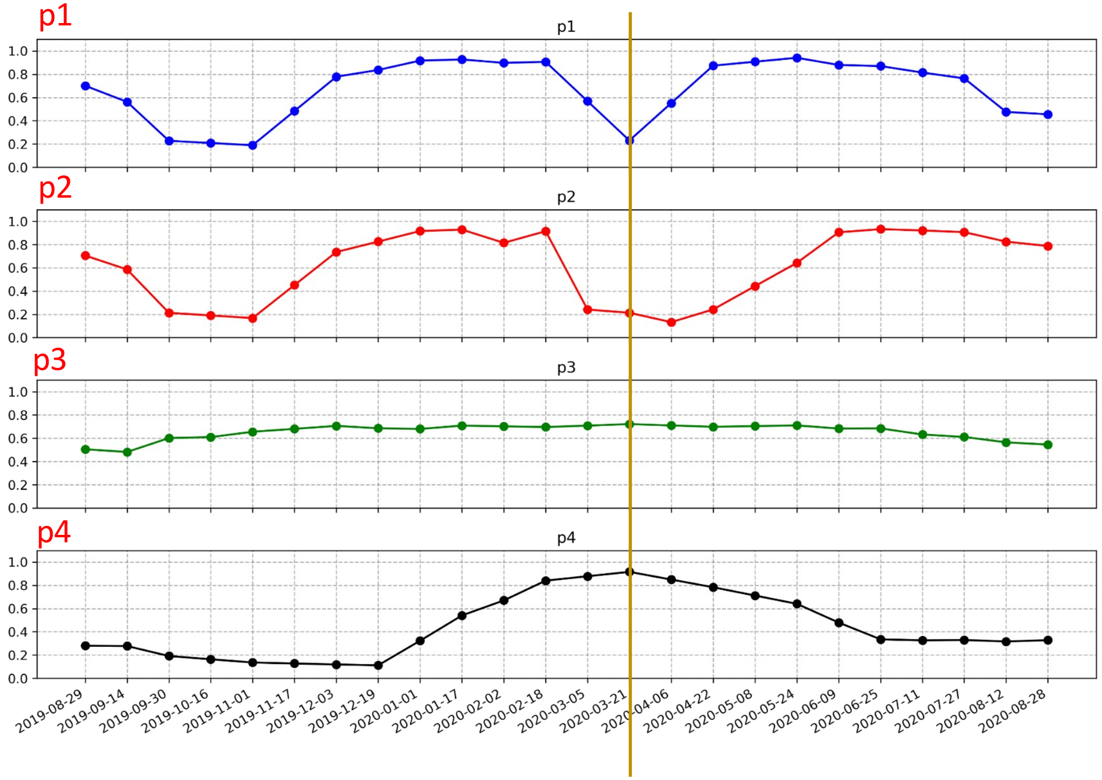
NDVI Time Series – Sentinel-2 Data Cube (16-days)

BRAZIL
DATA CUBE

Sentinel-2 RGB



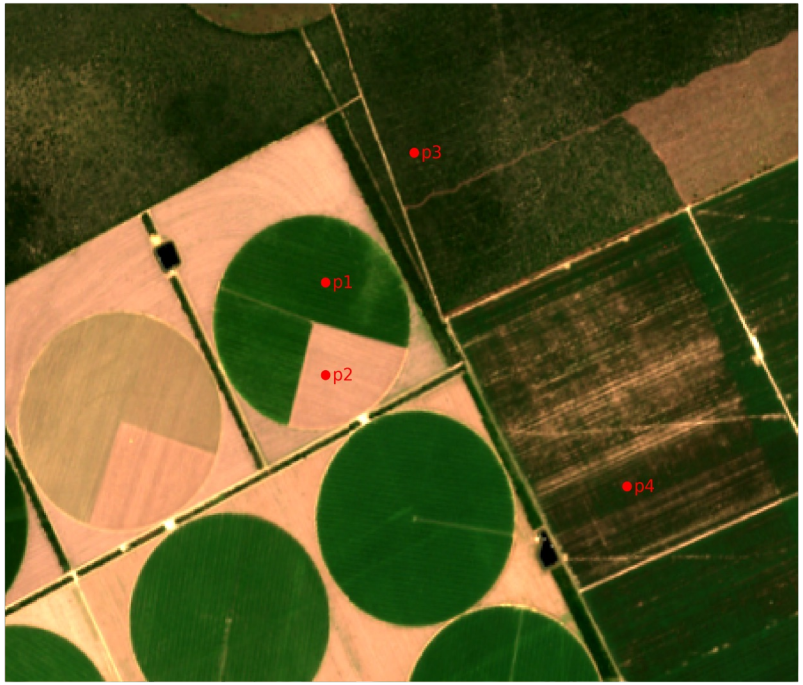
2020-03-21



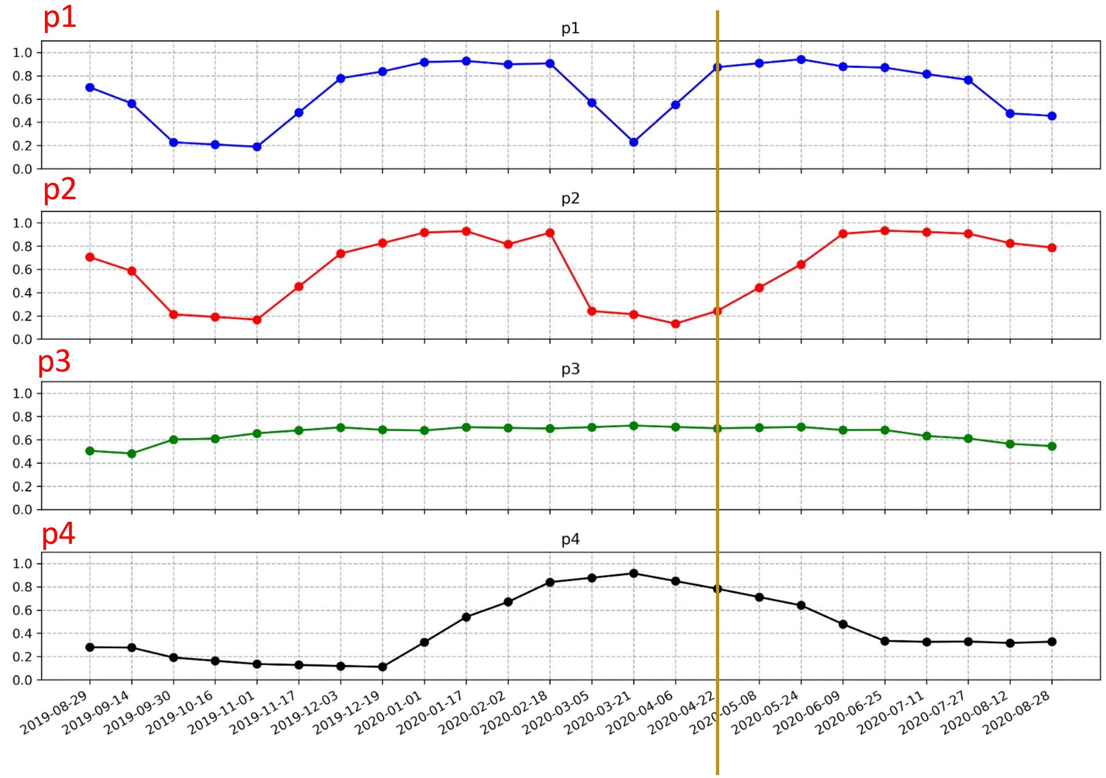
2020-03-21

NDVI Time Series – Sentinel-2 Data Cube (16-days)

Sentinel-2 RGB



2020-04-22



2020-04-22

Big data of remote sensing images modeled as multidimensional data cubes

Land use and cover mapping

BRAZIL
DATA CUBE

Image time series analysis

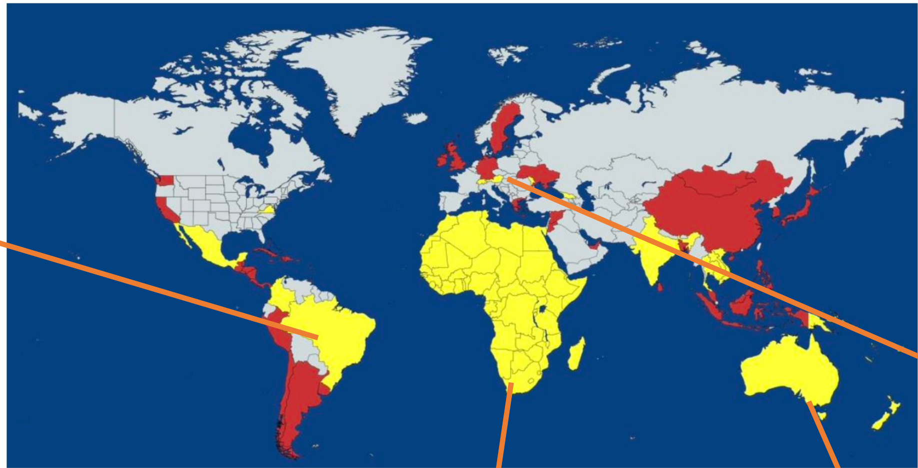
Big data technologies and machine learning

Research and technological innovation

Partnership with international and similar initiatives

Source: [Sudmanns et al, 2022]

Yellow: operational
Red: under development



BRAZIL
DATA CUBE

http://brazildatacube.org

data

Article

Paving the Way towards an Armenian Data Cube

https://www.swissdatacube.org/

BIG EARTH Data

Big Earth Data

Building an Earth Observations Data Cube: lessons learned from the Swiss Data Cube (SDC) on generating Analysis Ready Data (ARD)

ISSN: 2096-4471 (Print) 2574-5417 (Online) Journal homepage: http://www.tandfonline.com/loi/tbed20

Naciones Unidas
CEPAL

NASA

INEGI

Digital Earth Americas

A vision for a solution that provides relevant Earth observation data for the benefit of the Americas

Digital Earth
AFRICA

https://digitalearthafrica.org/

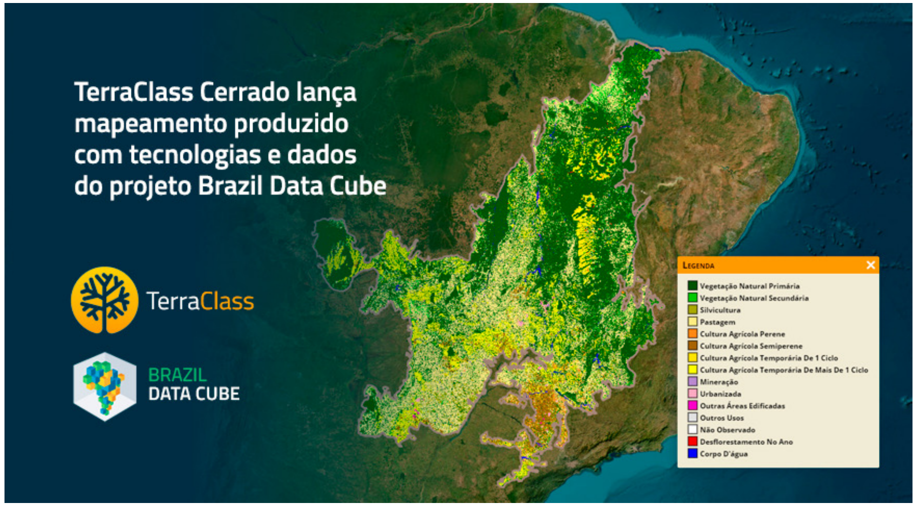
Digital Earth
AUSTRALIA

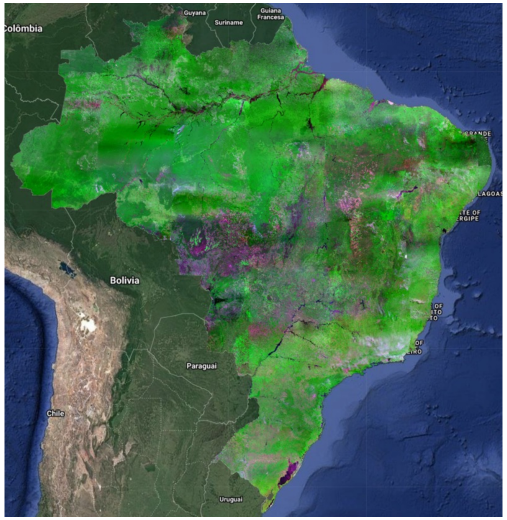
https://www.dea.ga.gov.au/

Motivation

Technological innovation for the environmental monitoring projects of INPE

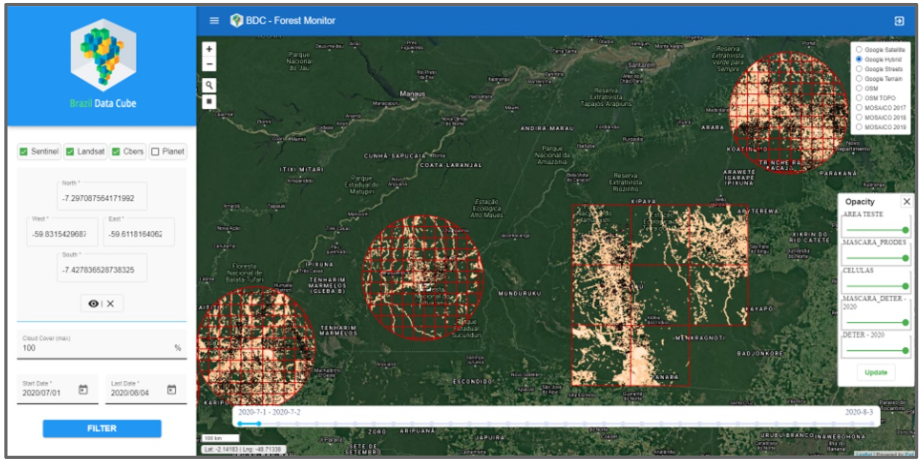
TerraClass Cerrado 2020 (launched in December 2022) using BDC data cubes and software tools





Mosaics – selection of the best pixels (free of clouds or cloud shadow) for periods.

Forest Monitor - **DETER Intenso**
Service to visualize big Earth observation data on AWS



Motivation

Image time series analysis and machine learning to produce land use and cover information from big Earth observation data

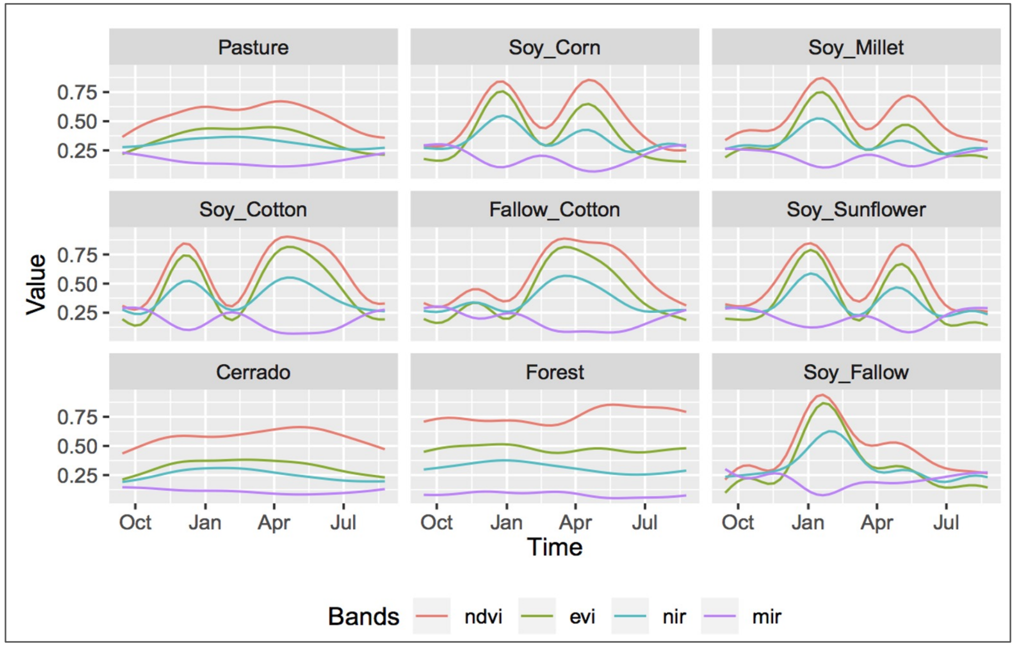
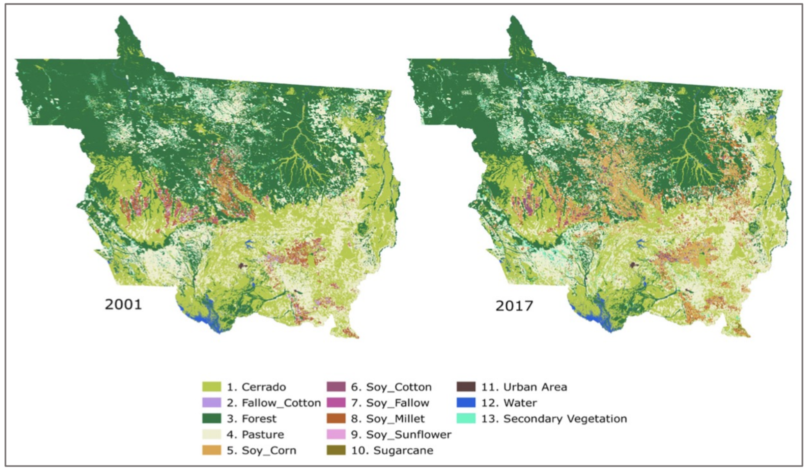


Image time series NDVI, EVI, NIR, MIR - agriculture year
MODIS – MOD13Q1 Product / Method – SVM (Support Vector Machine)



Land use and cover maps for Mato Grosso State in Brazil from 2001 to 2017, Scientific Data, 2020 (Simoes et al., 2020)

Motivation

Image time series analysis to extract vegetation phenological metrics.

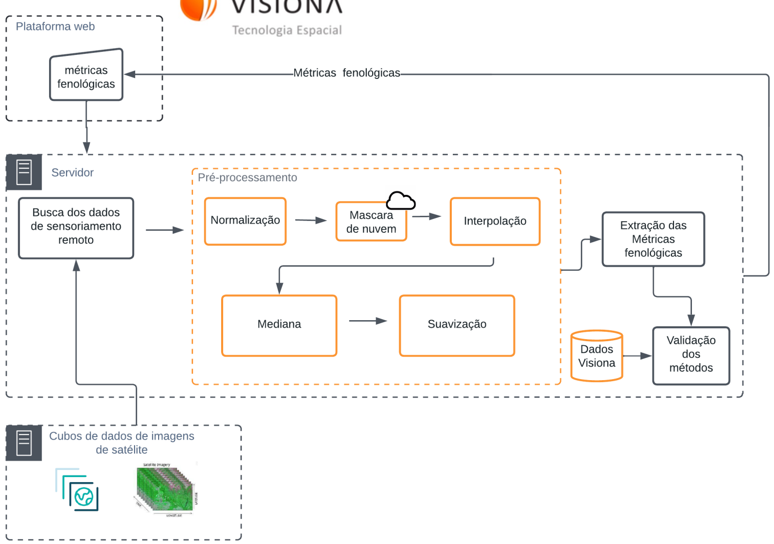
A review of vegetation phenological metrics extraction using time-series, multispectral satellite data, Remote Sensing of Environment, 2020 (Zeng et al., 2020)

Table 4
Summary of main phenological metrics extraction methods for species-specific vegetation types from satellite imagery.

Methods	Vegetation types	Sensors	Stage classification	Specific Stages	Method Classification	Reference
Line segment fitted parameters and statistics	Quercus petraea, Fagus ylvatica L	AVHRR	Physiological-based phenological stages	Budburst, senescence	Empirical Statistics method	Duchemin et al. (1999)
Inflection points determined by derivative	Rice	MODIS	Physiological-based phenological stages	Planting, heading, and harvesting	Empirical method	Sakamoto et al. (2005)
Based on the parameters derived from the best fitted polynomial curve	Potato	MODIS	General phenological stages	12 metrics for potato	Empirical method	Islam and Bala (2008)
Use TIMESAT software to detect rice phenological stages	Rice	MODIS	General phenological stages	Start, peak and end of season	Empirical method	Boschetti et al. (2009)
Derive phenological dates based on the optimum scaling parameters and shape model.	Corn and soybeans	MODIS	Physiological-based phenological stages	8 stages for corn and soybeans respectively	Phenology matching	Sakamoto et al. (2010)
Regress the ground measure degree days and VI values	Sugarcane	ASTER	Physiological-based phenological stages	6 stages	Simulation	Mobasheri et al. (2010)

Motivation

Image time series analysis to extract vegetation phenological metrics.



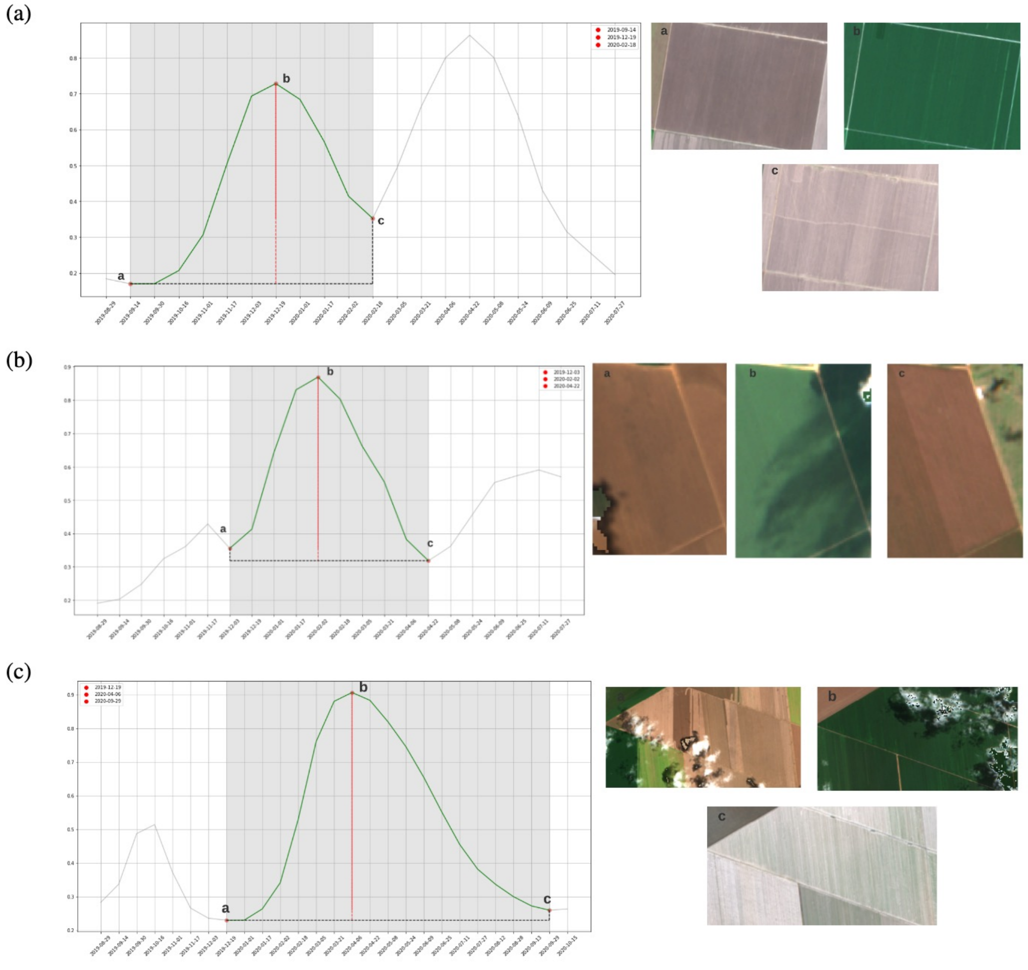


Figura 4. Métricas fenológicas de início, fim e máximo vigor vegetativo de plantio extraídas para soja (a), milho de primeira safra (b) e algodão (c) utilizando séries Sentinel-2A/B de NDVI a cada 16 dias.

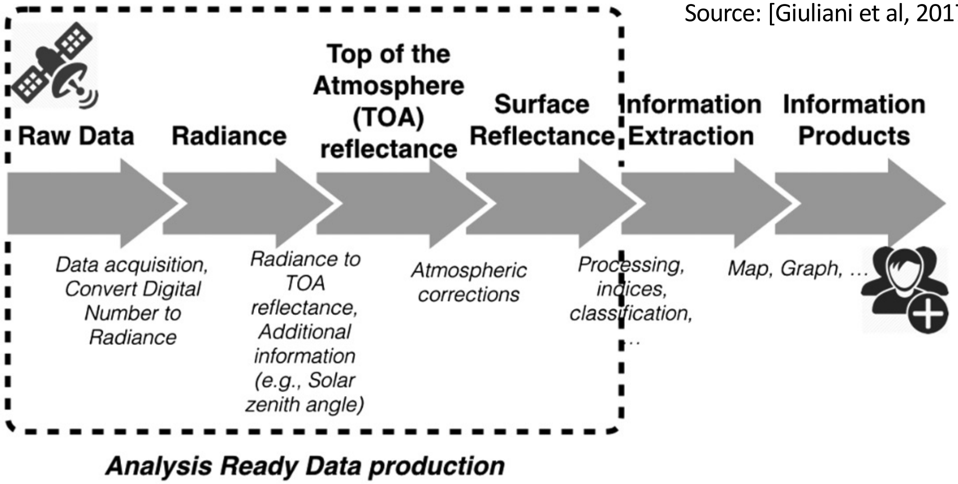
Objectives

(1) Analysis-Ready Data (ARD) of medium-resolution satellite images for Brazil: CBERS-4 Landsat 8 Sentinel 2.

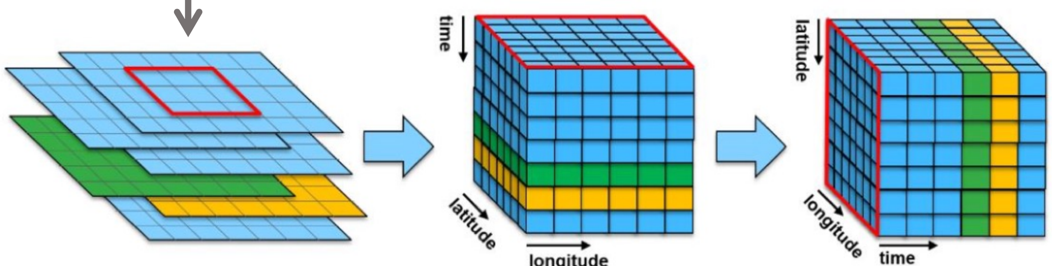
(2) Multidimensional data cubes.

CEOS Analysis Ready Data for Land:
https://ceos.org/ard/index.html

Source: [Giuliani et al, 2017]



Source: [Kopp et al, 2019]



Collections of Earth observation satellite imagens – ARD

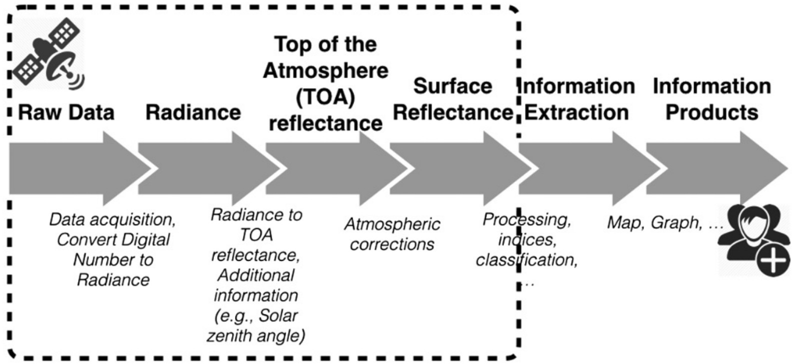
Data cubes – four-dimensional array.

ARD and EO Data Cubes for the entire Brazilian territory

BRAZIL
DATA CUBE

2 Petabytes (PB) remote sensing images

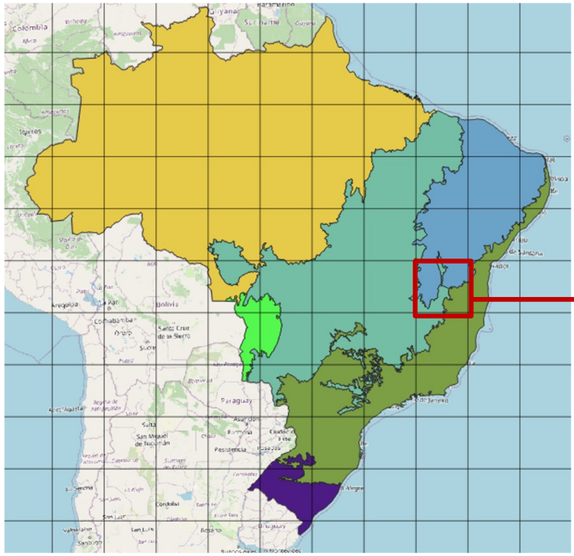
Sentinel-2
Landsat-8 / -9
CBERS-4 / 4A (MUX / WFI)



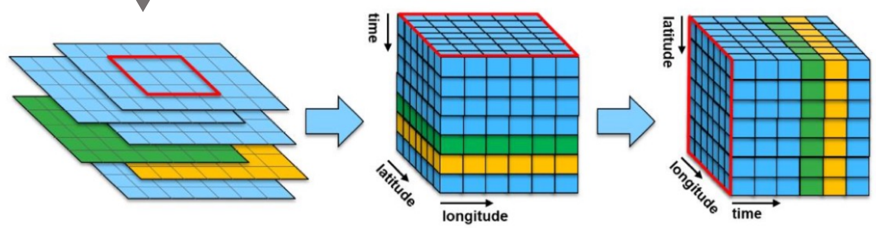
Analysis Ready Data production

Source: [Giuliani et al, 2017]

CEOS Analysis Ready Data (ARD) for Land:
https://ceos.org/ard/index.html



BDC Hierarchical Tiling System



Collections of ARD satellite imagens

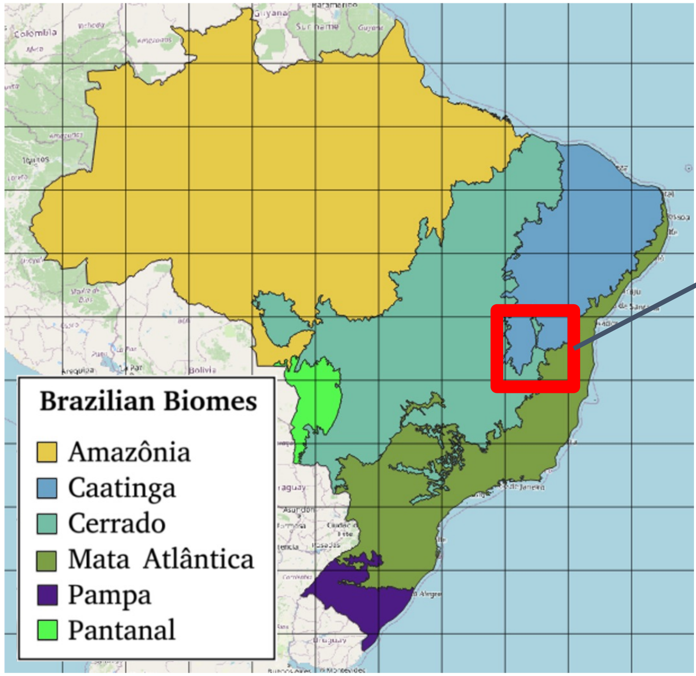
Multidimensional Data cubes

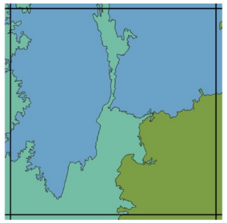
Source: [Kopp et al, 2019]

BRAZIL
DATA CUBE

Hierarchical tiling system – 3 Grids

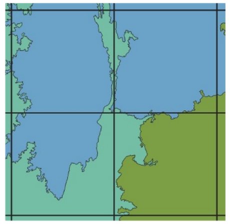
BDC Grid (V2)





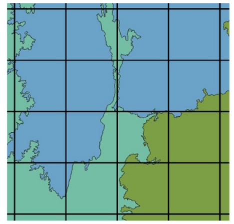
Grid: BDC – Large
Tile size: 4224400m x 4224400m

422.4 km



Grid: BDC – Medium
Tile size: 211200m x 211200m

211.2 km



Grid: BDC – Small
Each tile: 105600m x 105600m

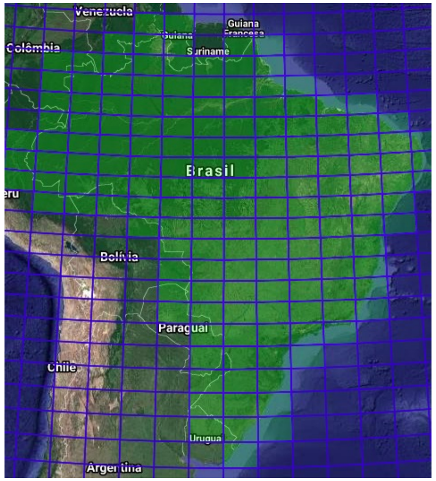
105.6 km

Projection: Albers equal area and Datum: SIRGAS 2000 brazildatacube.dpi.inpe.br/portal/explore

Temporal-composed Data Cubes

CBERS-4/MUX (20 meters)

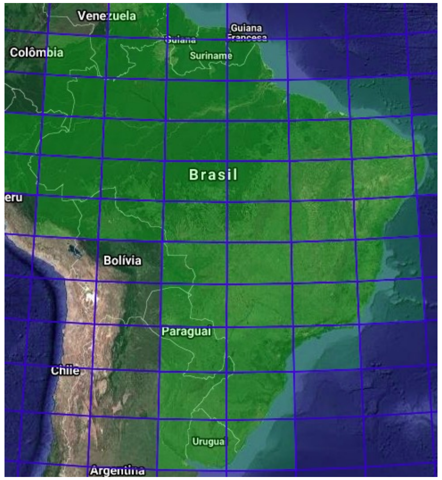
Data cubes stack – 1 Month
[2016, 2022]



Grid: BDC – Medium
Each tile: 3 x 2 degrees

CBERS-4/AWFI (64 meters)

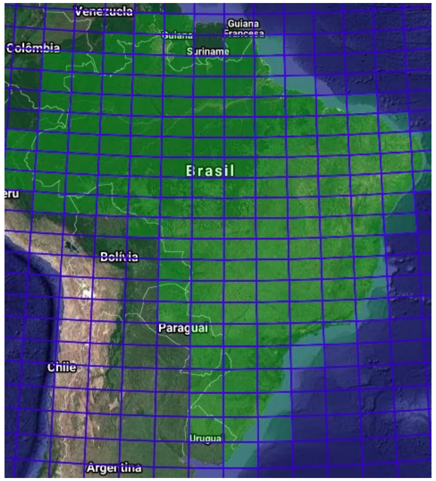
Data cubes stack – 16 days
[2016, 2022]



Grid: BDC – Large
Each tile: 6 x 4 degrees

Landsat-8/OLI (30 meters)

Data cubes stack – 16 days
[08/2016 to 10/2021]



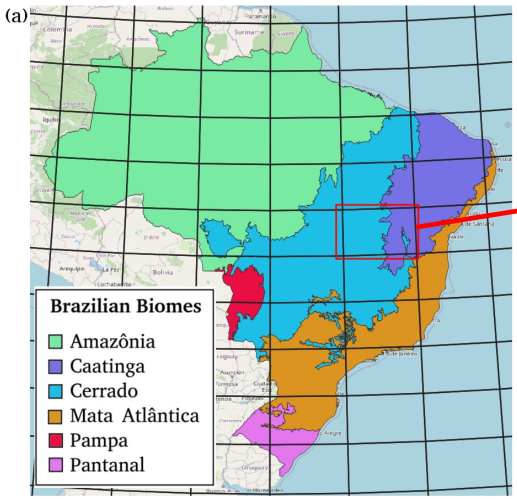
Grid: BDC – Medium
Each tile: 3 x 2 degrees

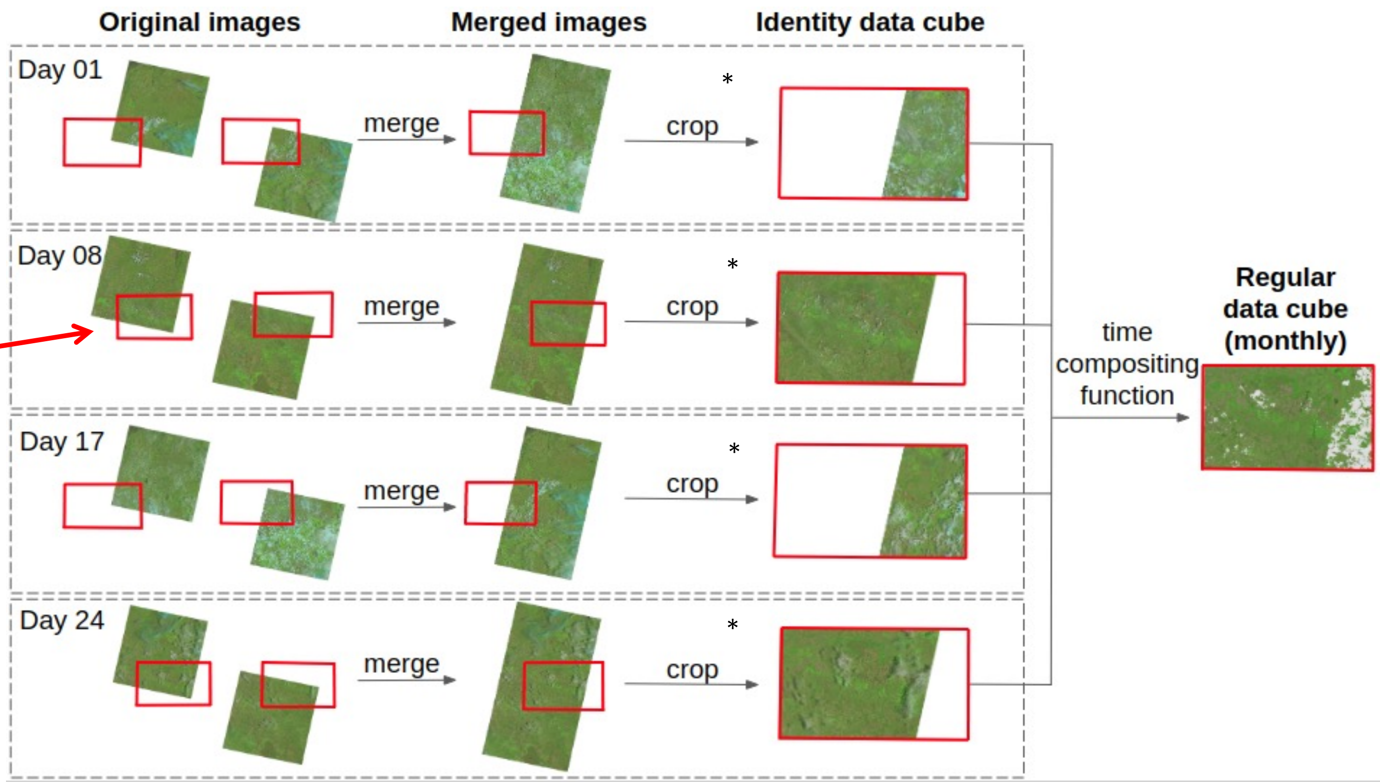
Sentinel-2/MSI (10 meters)

Data cubes stack – 16 days
[2017, 2020]

Grid: BDC – Small
Each tile: 1.5 x 1 degree

Building data cubes

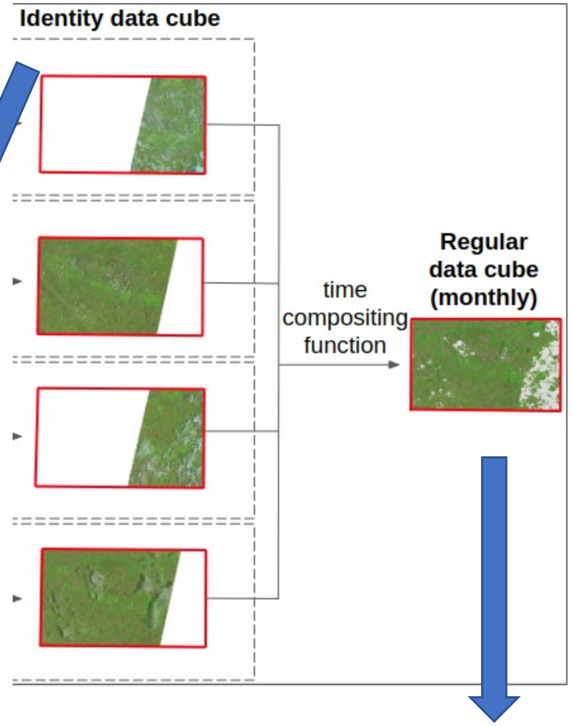




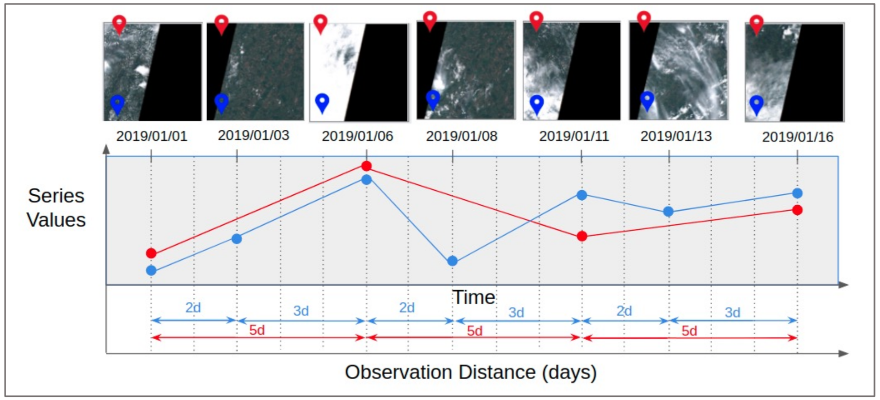
*bilinear resampling for better spatial resolution band

BRAZIL
DATA CUBE

Identity and **Temporal-composed** data cubes



Identity data cubes: produced using all available images in a time interval (ex. a month or 16 days). Time series extracted from them can be or not regular in time.

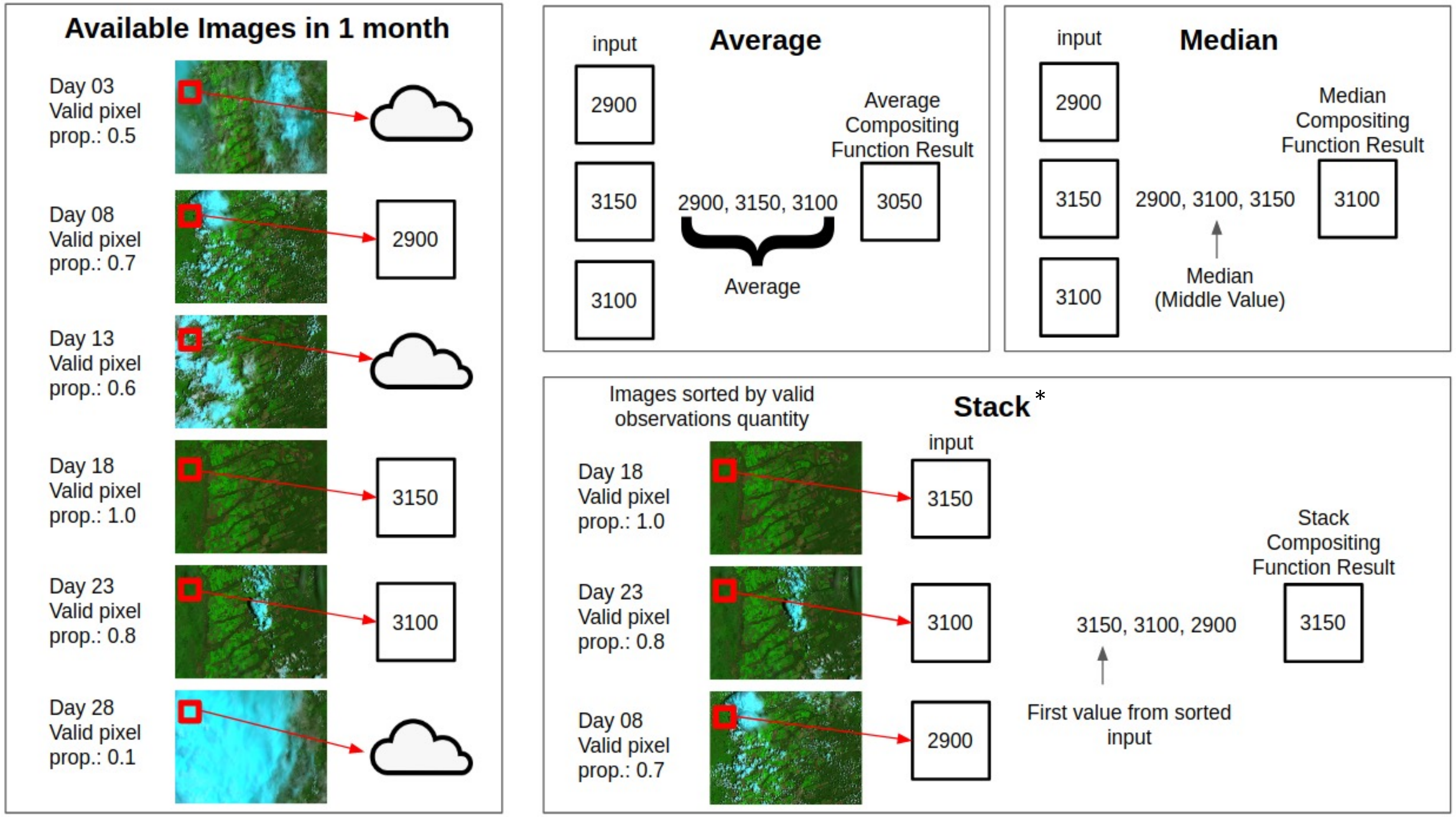


Time series extracted from identity EO data cubes can be regular in time (red time series) or not (blue time series)

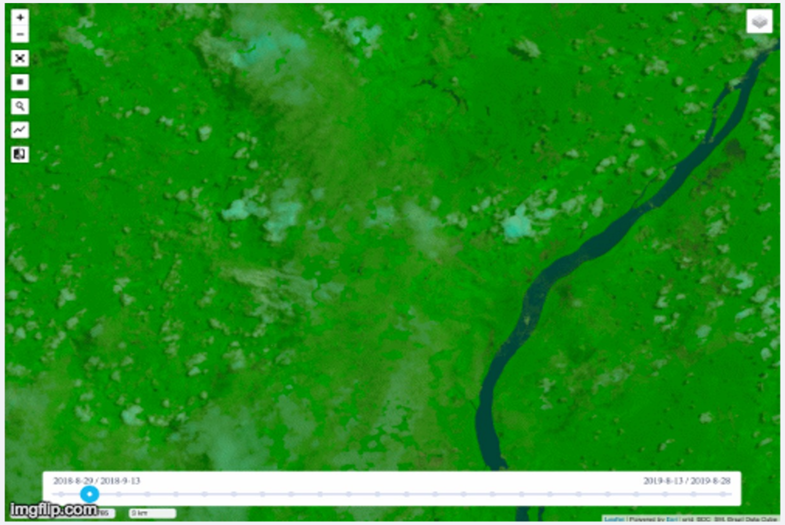
Temporal-composed data cubes: produced using a temporal compositing function to select the best pixels (free of cloud and cloud shadow) obtained in each period (ex. a month or 16 days). Time series extracted from them are always regular in time.

Building data cubes

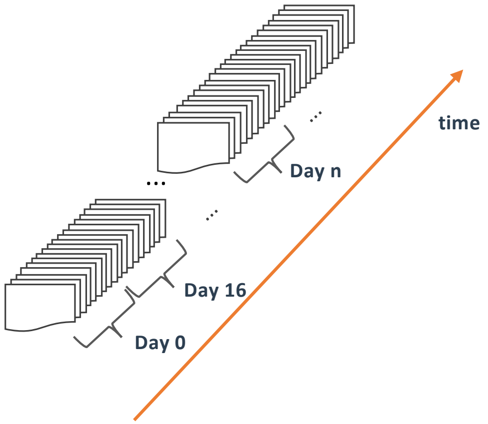
Time Compositing



* Least Cloud Cover First



brazildatacube.dpi.inpe.br/portal/explore



For each *tile* and *time step*, there are a set of COG (Clould Optimized GeoTIFF) files:

(1) Spectral bands from original images; (2) Vegetation indices (EVI and NDVI); (3) Cloud mask; (4) number of valid observations (excluding cloud, cloud shadow..); (5) data provenace; ...

BDC – Small
Each tile: 105600m x 105600m

Sentinel-2/MSI – 10 meters
Each file (band/tile): 400 MB
Each tile: ~ 5.4 GB

BRAZIL
DATA CUBE

Mosaics

Landsat-8 - OLI - Brazil

6 months
[from July to December - 2017]

RGB: B6 -B5 -B4
Resolution: 30m

CBERS-4 WFI - Brazil

3 months
[from May to June - 2020]

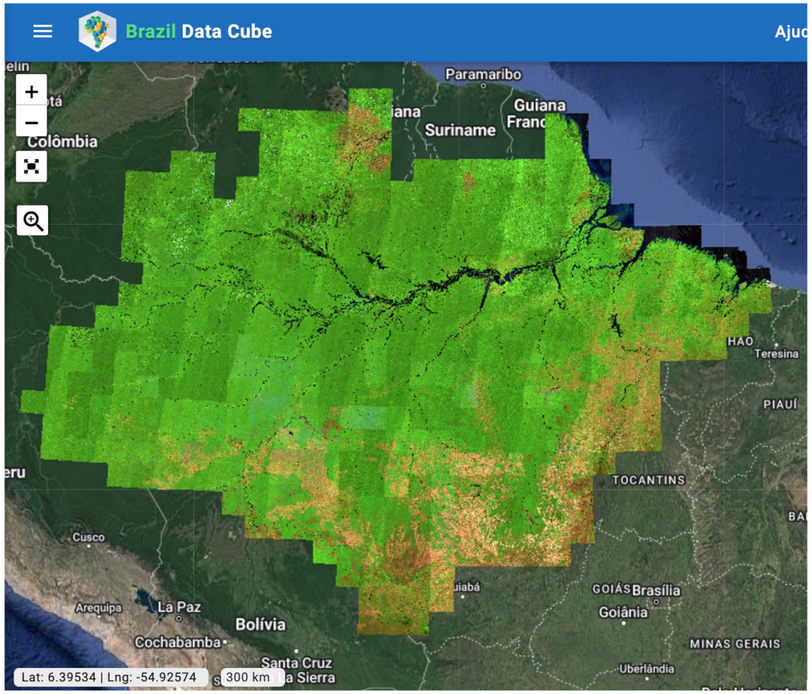
RGB: B15 - B16 - B13
Resolution: 64m

BRAZIL
DATA CUBE

Mosaics

Sentinel-2 - MSI - Amazonia

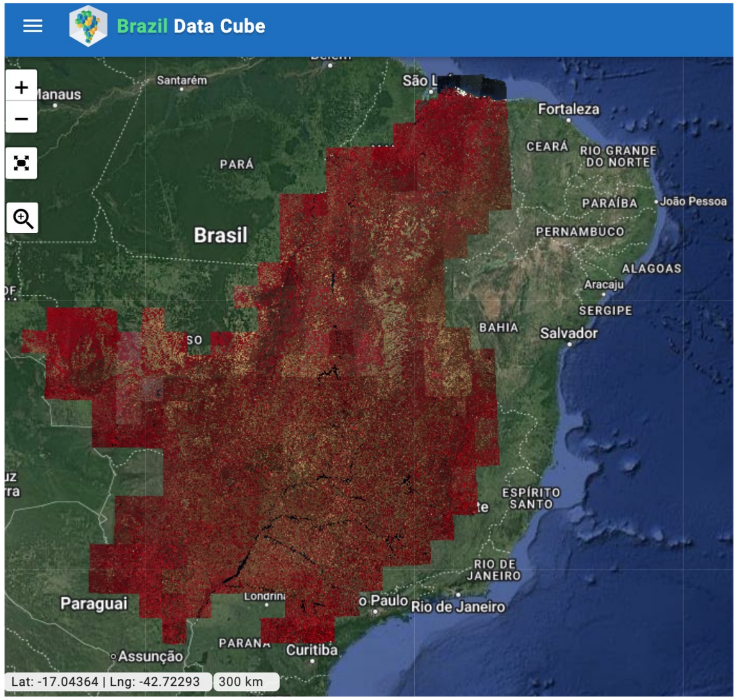
3 months
[from June to August - 2022]



RGB: B11,B8A,B04
Resolution: 10m

Sentinel-2 - MSI - Cerrado

4 months
[from June to September - 2022]

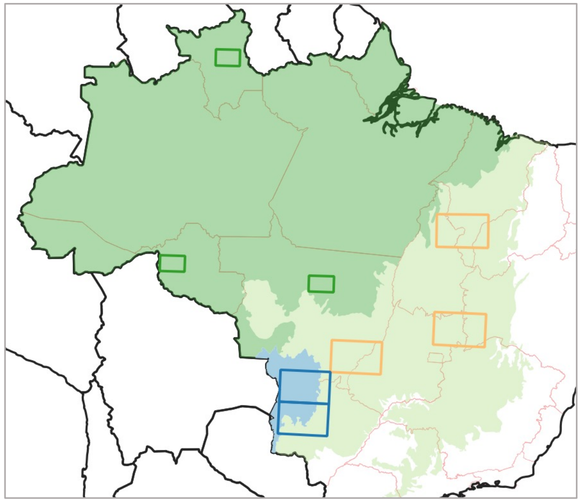


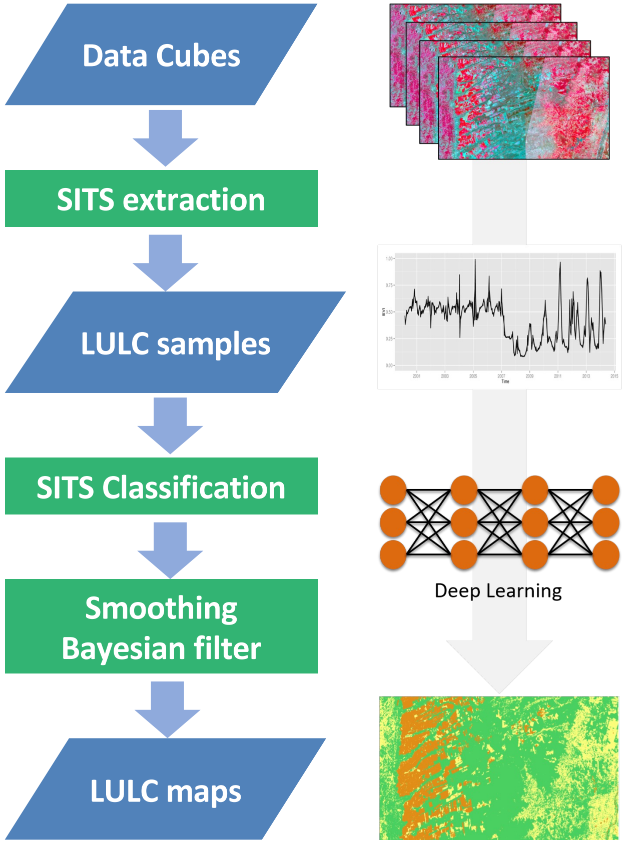
RGB: B08,B04,B03
Resolution: 10m

Objective

(3) Big data technologies, image time series analysis and machine learning methods .

(4) Land use and cover classification.

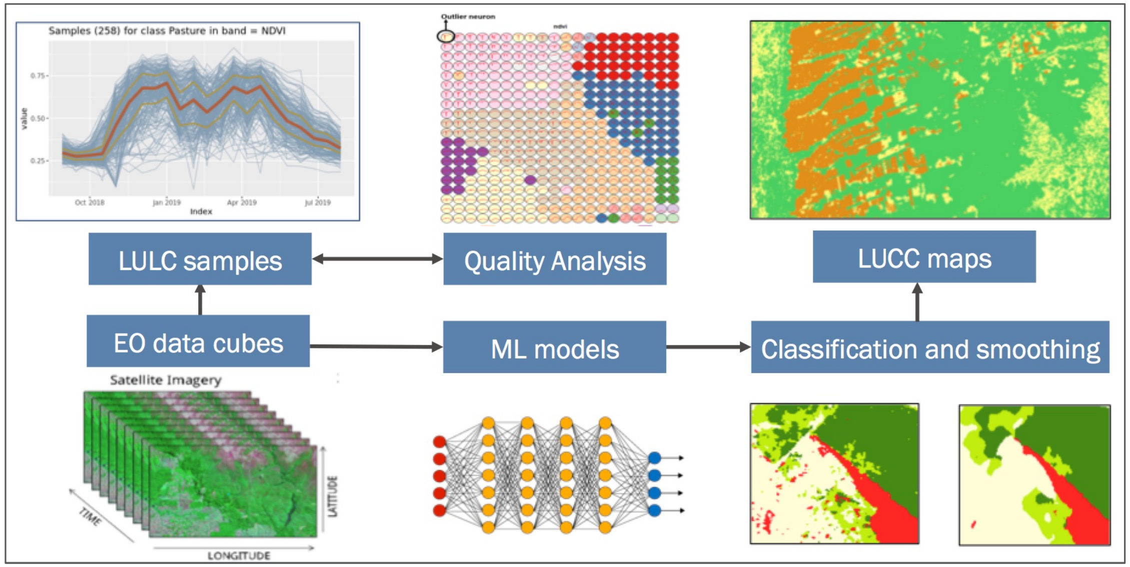




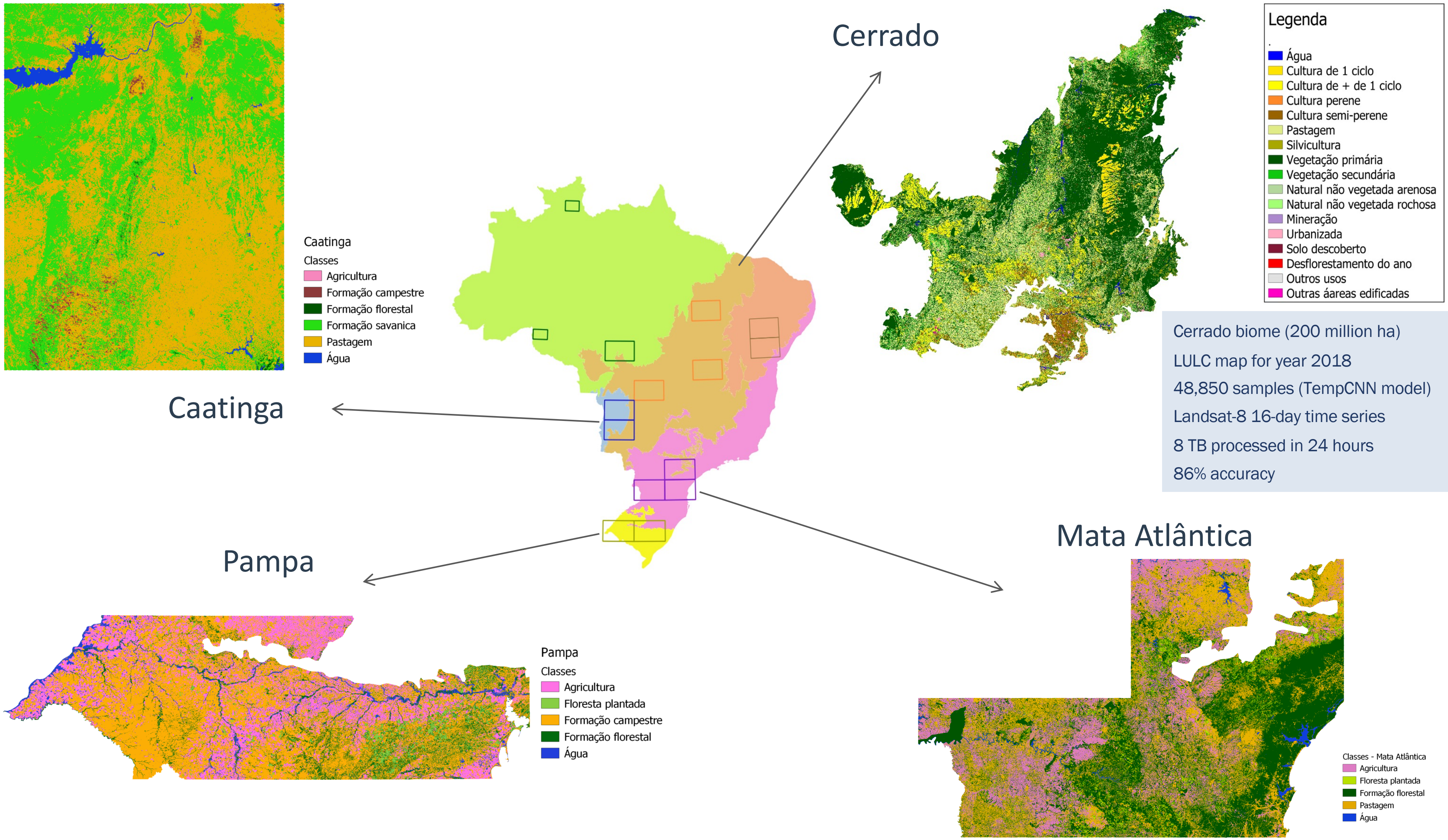
Source: [Ferreira et al, 2020]

SITS (Satellite Image Time Series) R package

https://github.com/e-sensing



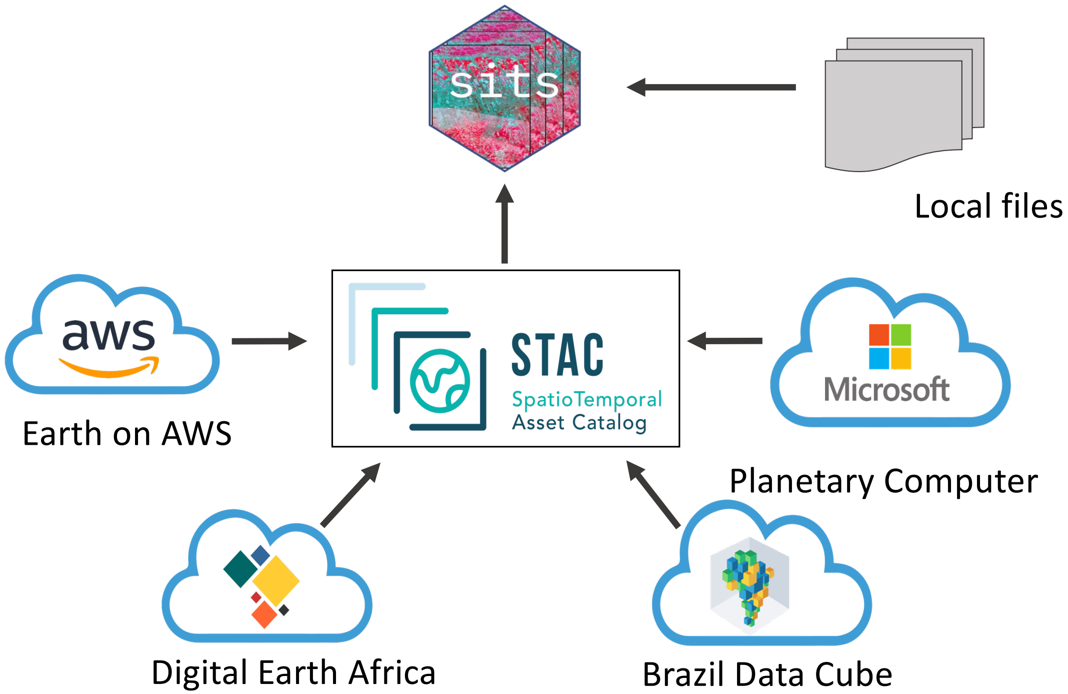

```r
cube <- sits_cube(source = "BDC",
                  collection = "S2_10_16D",
                  name = "s2_cube",
                  bands = c("NDVI", "EVI"),
                  tiles = "022024", start_date = "2018-09-01", end_date = "2019-08-28")

samples <- readRDS(url("https://bdc.dpi.inpe.br/rds/S2_10_16D.rds"))

mlp_model <- sits_deeplearning(layers = c(512, 512, 512, 512, 512), activation = "relu")

dl_model <- sits_train(samples, mlp_model)

result <- sits_classify(data = cube, ml_model = dl_model, output_dir = tempdir())
```



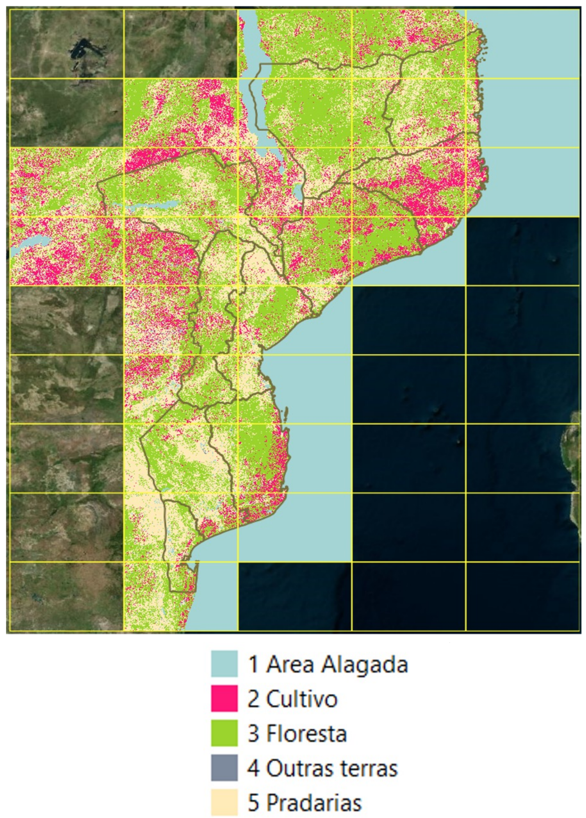
Cerrado
Legenda
Água
Cultura de 1 ciclo
Cultura de + de 1 ciclo
Cultura perene
Cultura semi-perene
Pastagem
Silvicultura
Vegetação primária
Vegetação secundária
Natural não vegetada arenosa
Natural não vegetada rochosa
Mineração
Urbanizada
Solo descoberto
Desflorestamento do ano
Outros usos
Outras áareas edificadas
Cerrado biome (200 million ha)
LULC map for year 2018
48,850 samples (TempCNN model)
Landsat-8 16-day time series
8 TB processed in 24 hours
86% accuracy
Caatinga
Classes
Agricultura
Formação campestre
Formação florestal
Formação savanica
Pastagem
Água
Caatinga
Pampa
Pampa
Classes
Agricultura
Floresta plantada
Formação campestre
Formação florestal
Água
Mata Atlântica
Classes - Mata Atlântica
Agricultura
Floresta plantada
Formação florestal
Pastagem
Água

SITS (Satellite Image Time Series) R package

https://github.com/e-sensing

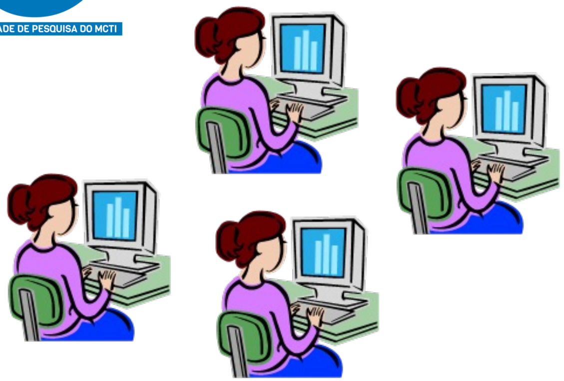


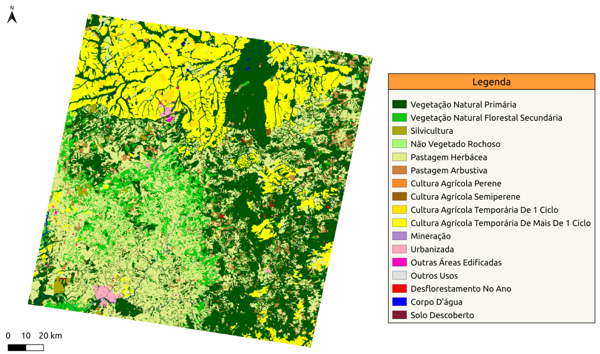
Land use and cover map for Mozambique - 2016

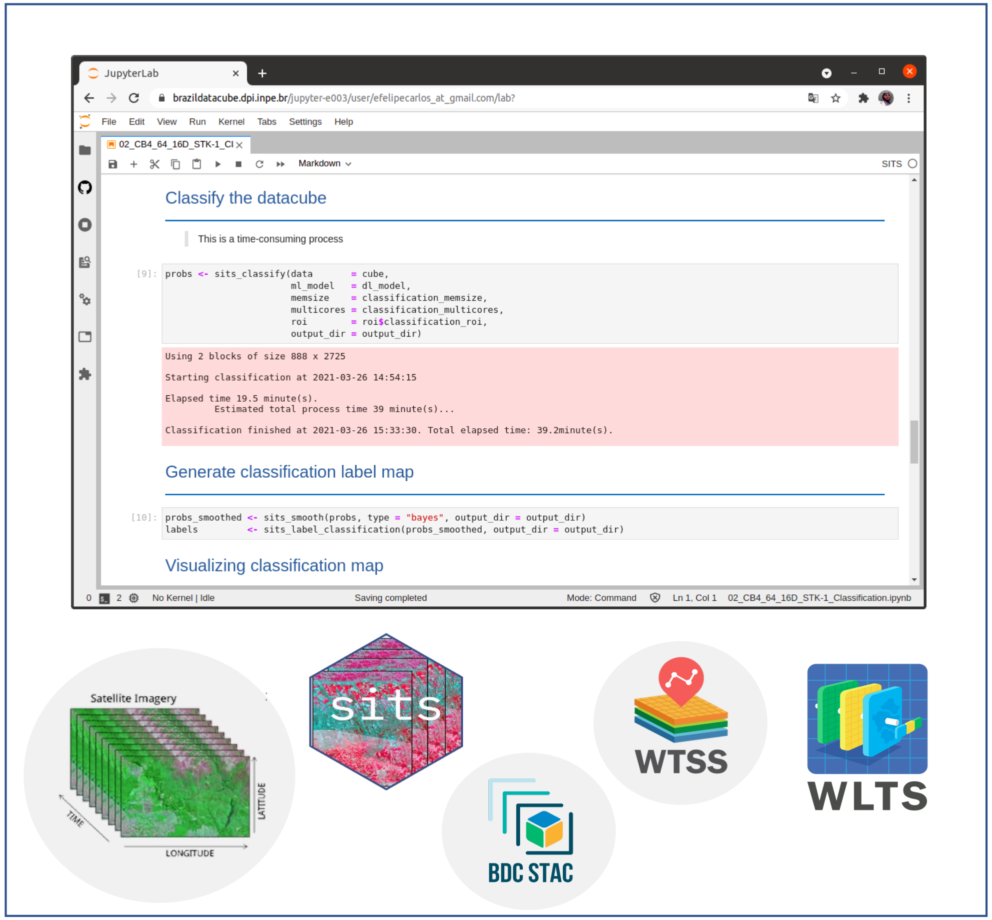


BRAZILDATACUBE.ORG

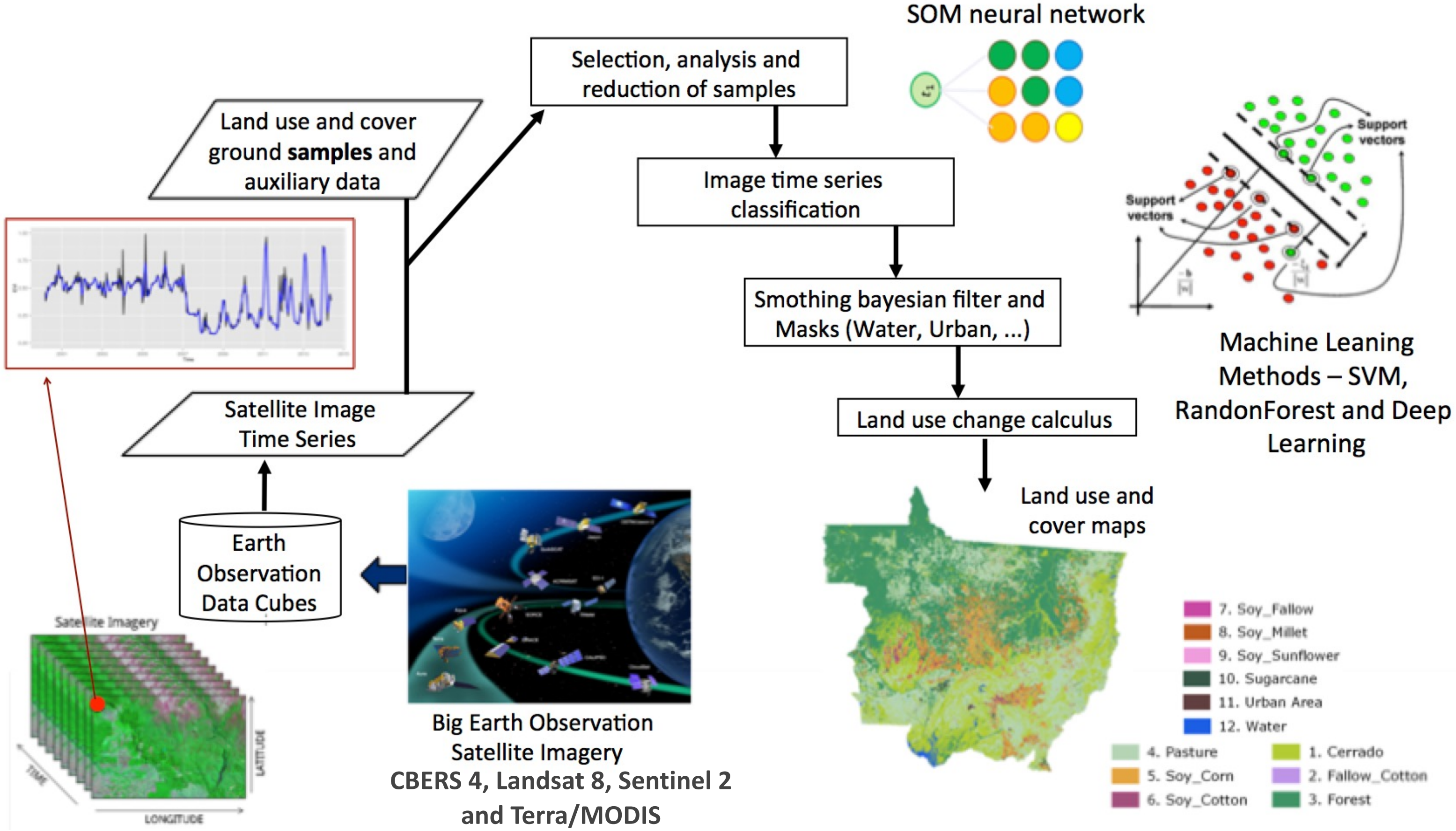
Brazil Data Cube technologies - INPE







SITS (Satellite Image Time Series) R package:
https://github.com/e-sensing

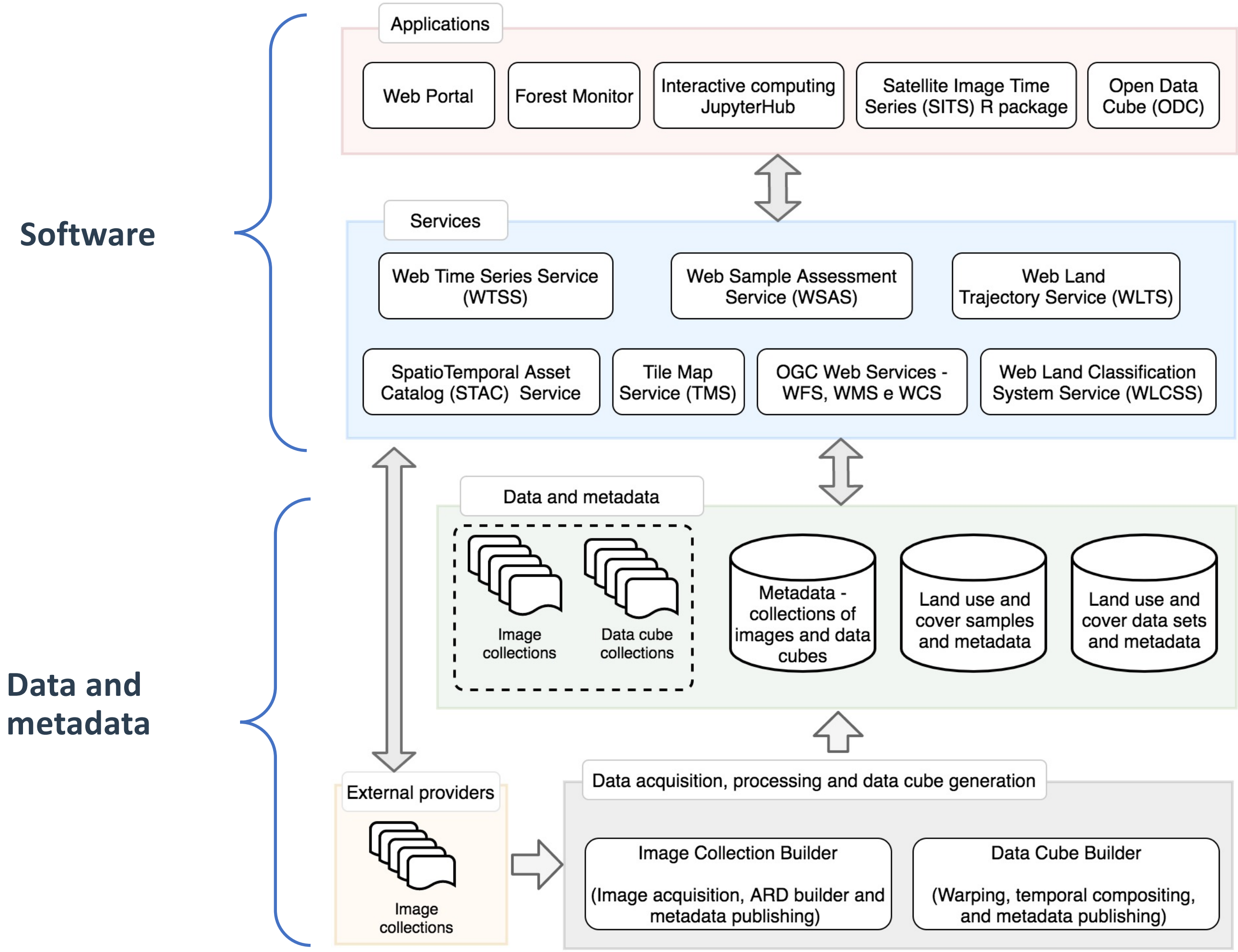


ARD and Data cubes available at:
http://brazildatacube.dpi.inpe.br/portal/explore

Land use and cover change maps:
http://brazildatacube.dpi.inpe.br/portal/explore

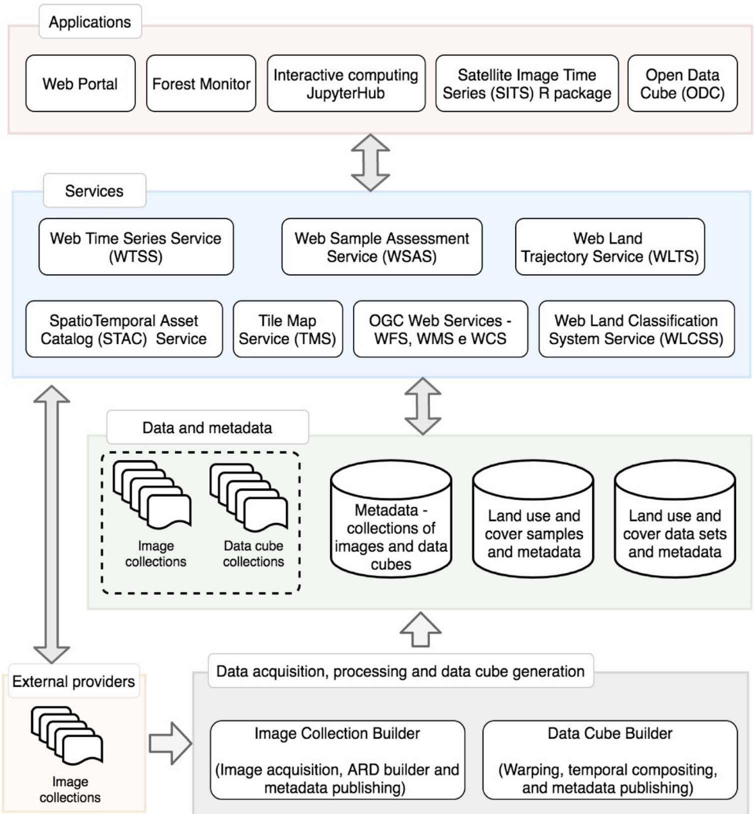
BRAZIL DATA CUBE

Open Data and Software Products

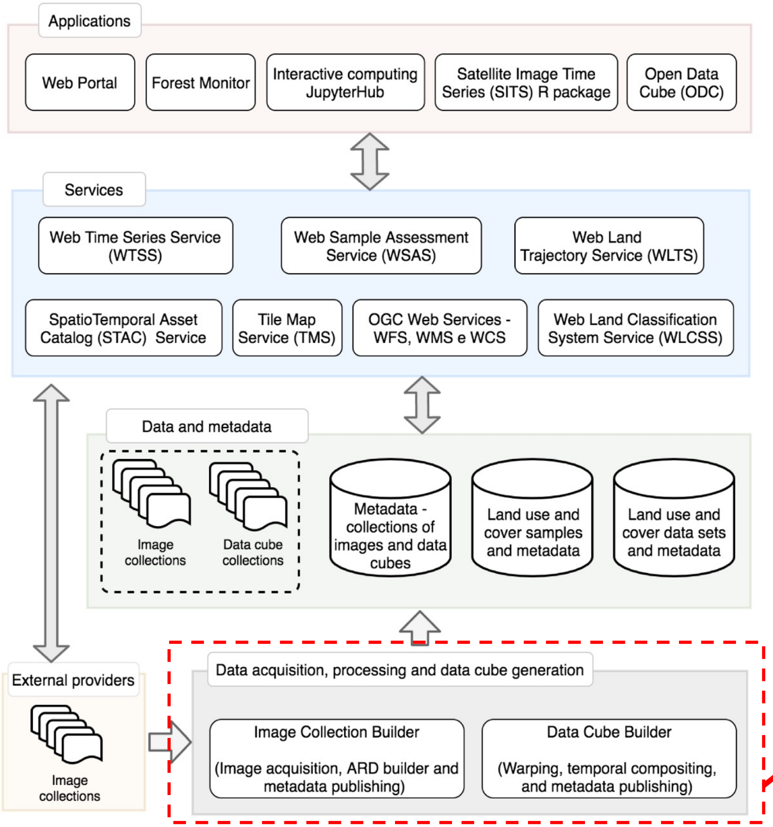


Source: [Ferreira et al., 2020]

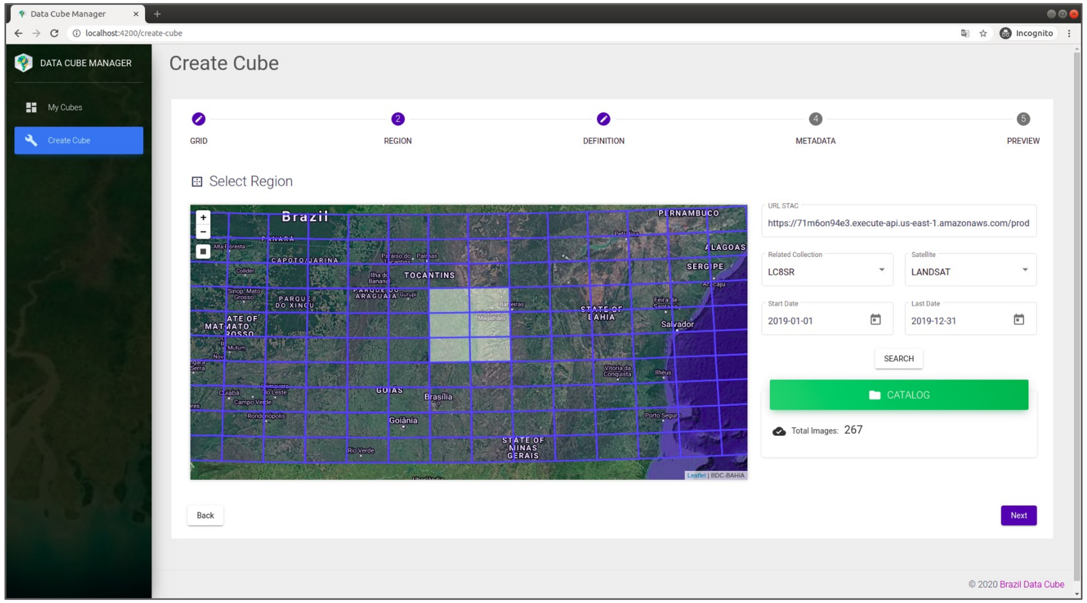
Computational Platform

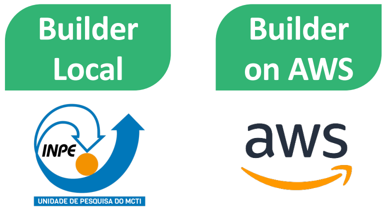


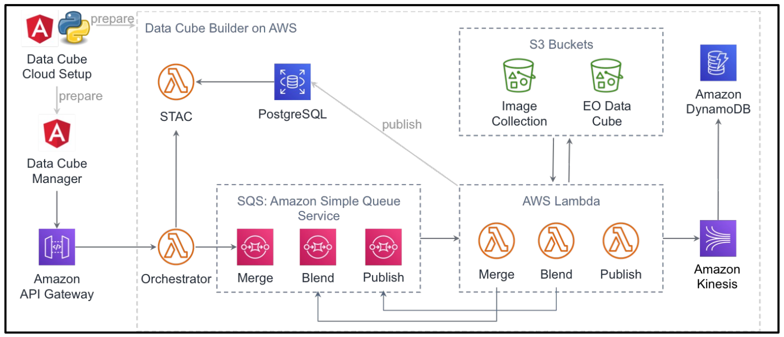
Source: [Ferreira et al, 2020]



Data Cube Builder







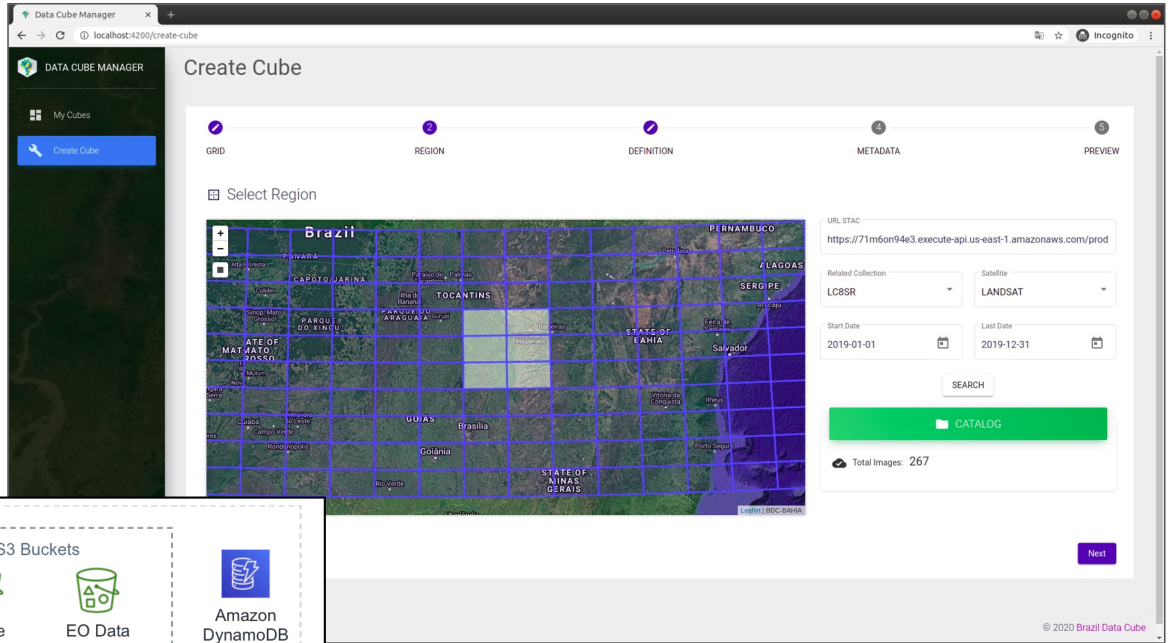
Source: [Ferreira et al, 2022]

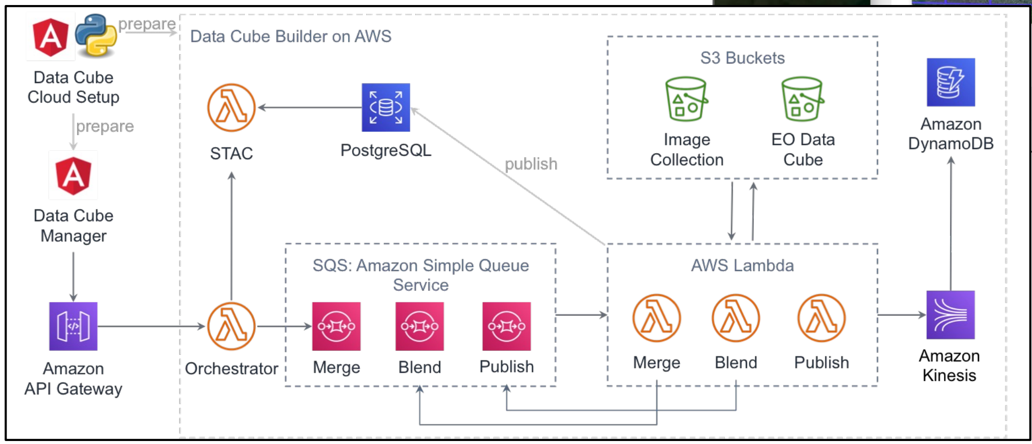
Amazon Web Services (AWS)

Build Sentinel-2 data cubes on AWS for 2021

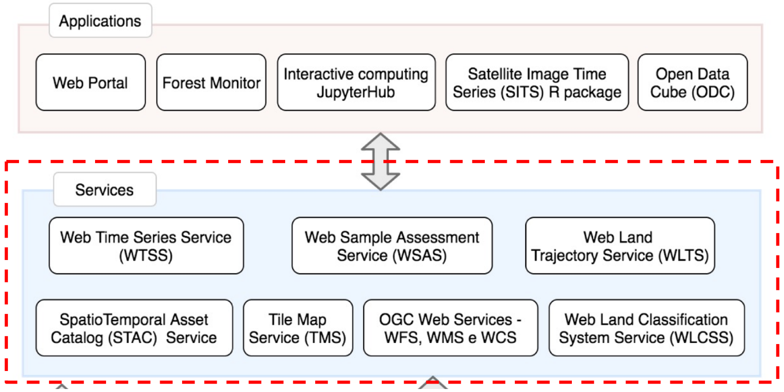
AWS Open Data

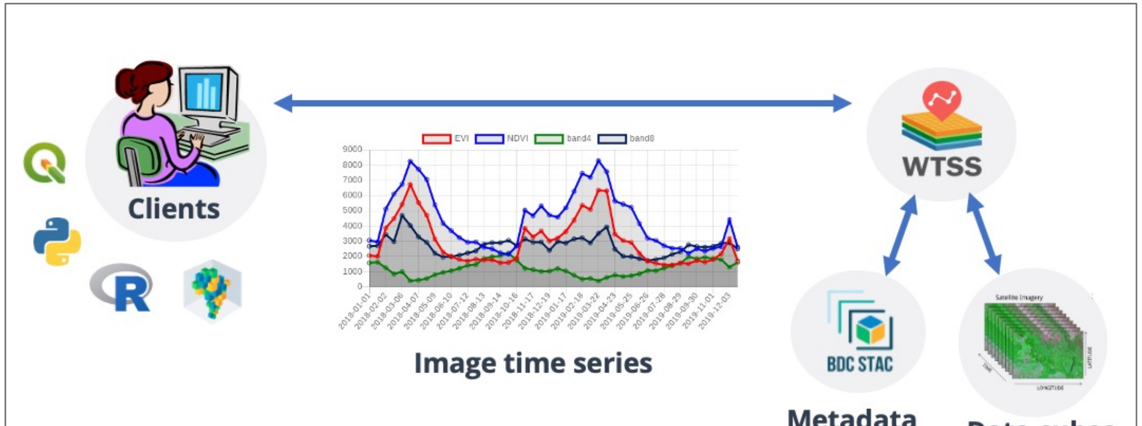
Data Cube Builder On AWS





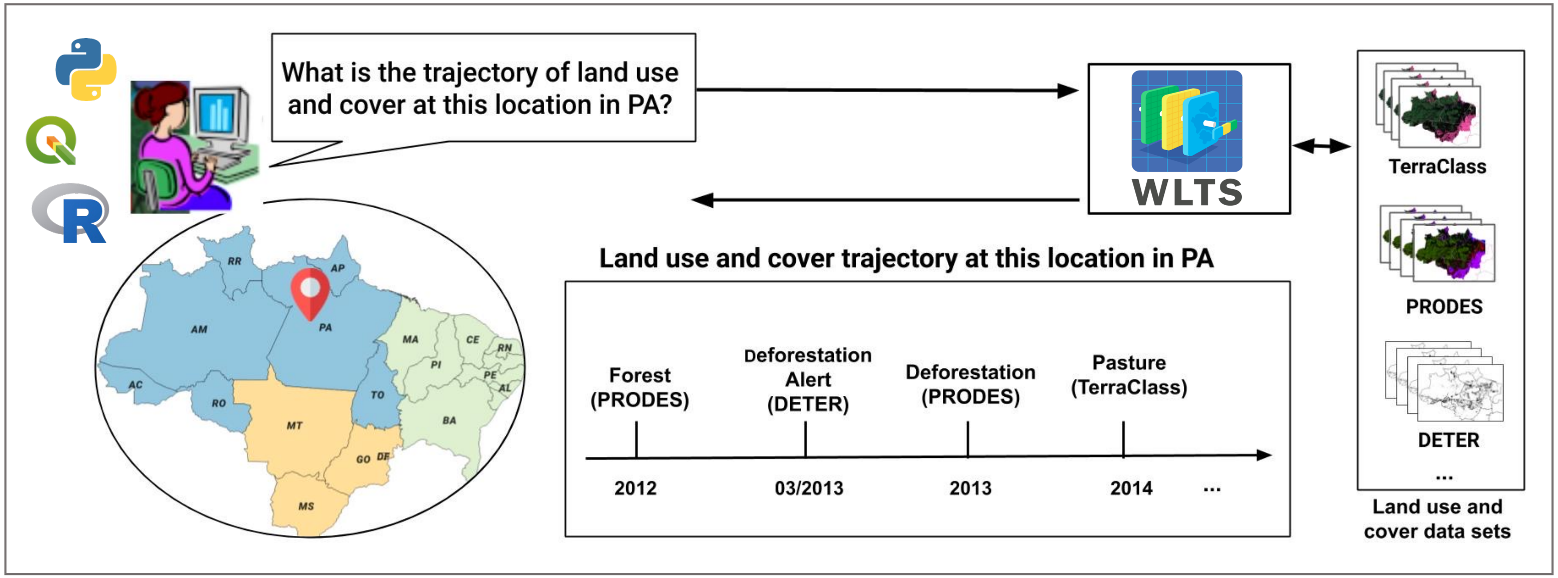
GEO AWS Cloud Credit Program
U$ 60,000





WTSS – Web Time Series Service

WLTS – Web Land Trajectory Service

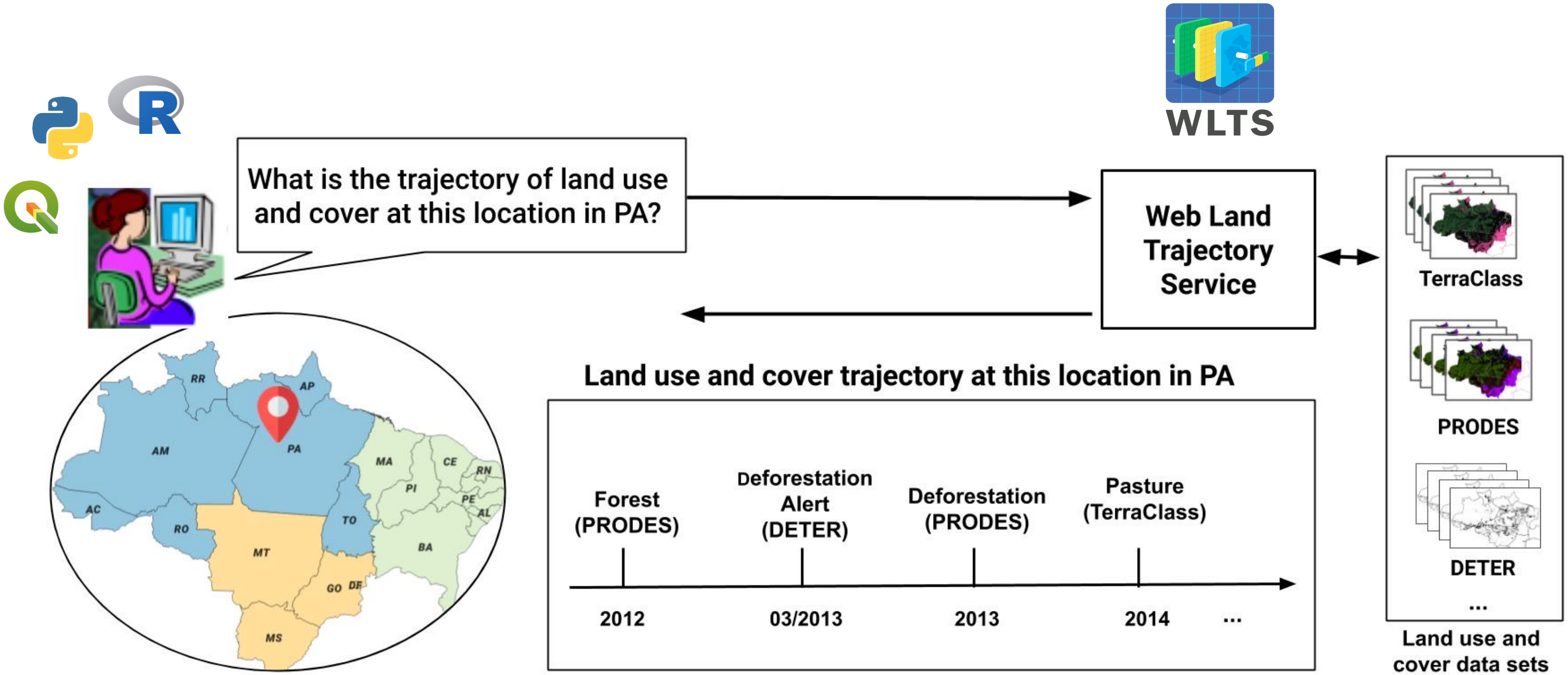


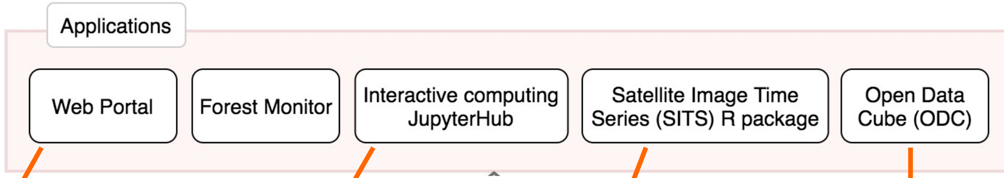
BDC STAC

Web Time Series Service (WTSS)

Image time series

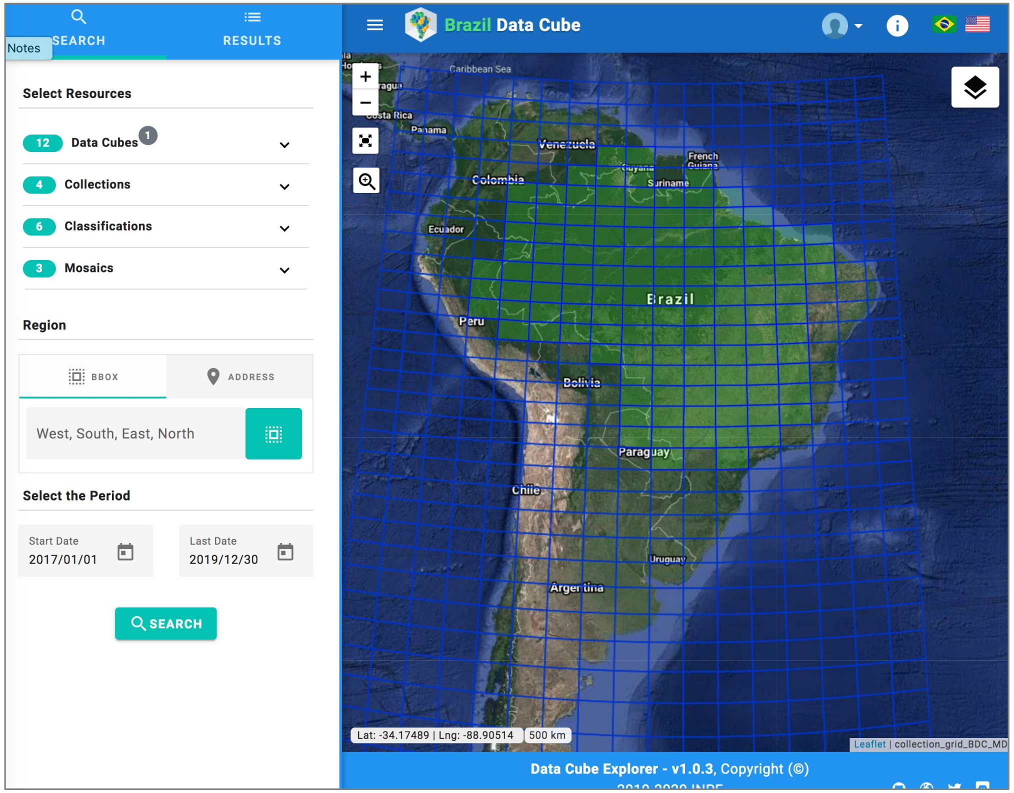
Web Land Trajectory Service (WLTS)





BDC Explorer 3.0

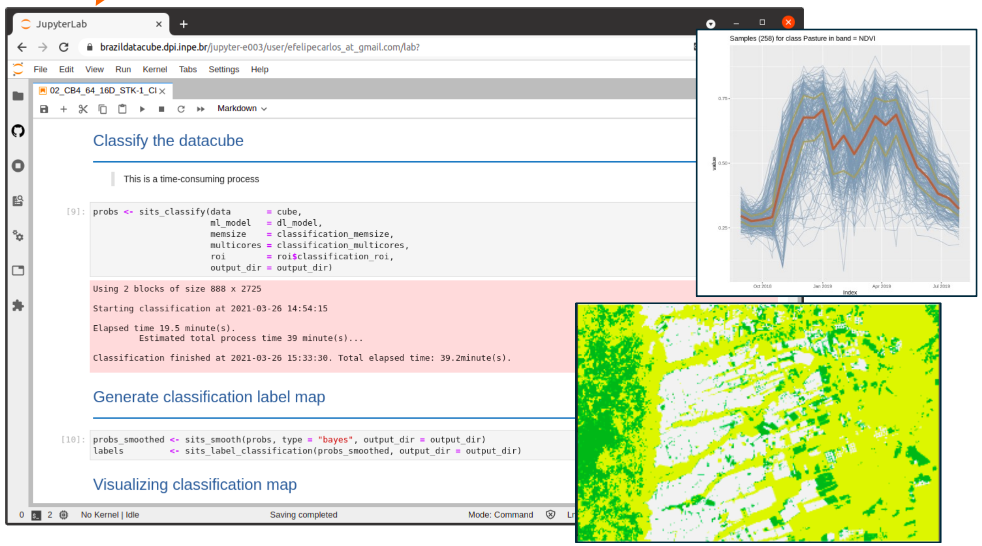
http://brazildatacube.dpi.inpe.br/portal/explore



SITS (Satellite Image Time Series) R package:

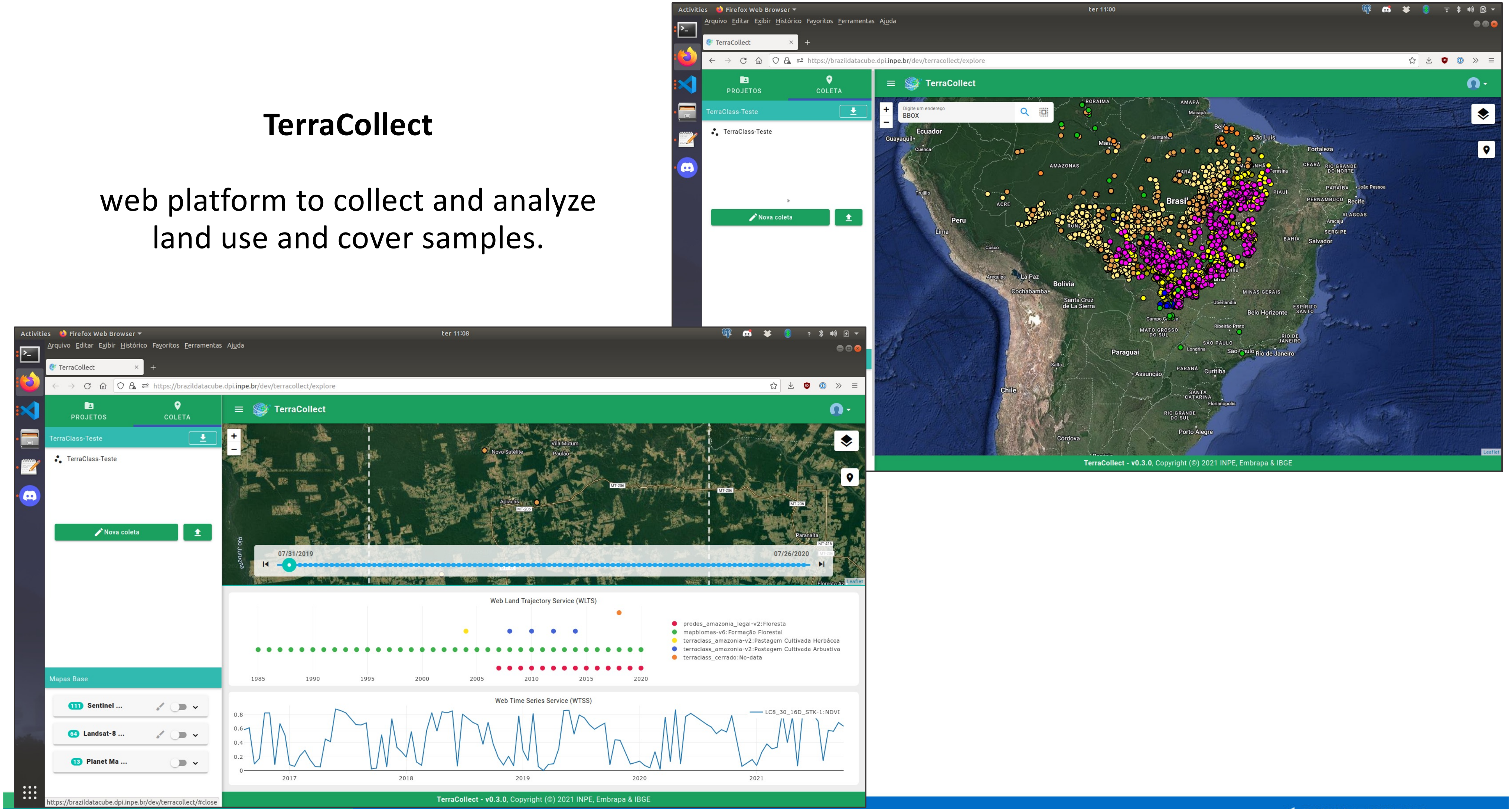
https://github.com/e-sensing

BDC JupyterHub : Interactive computing



TerraCollect

web platform to collect and analyze land use and cover samples.



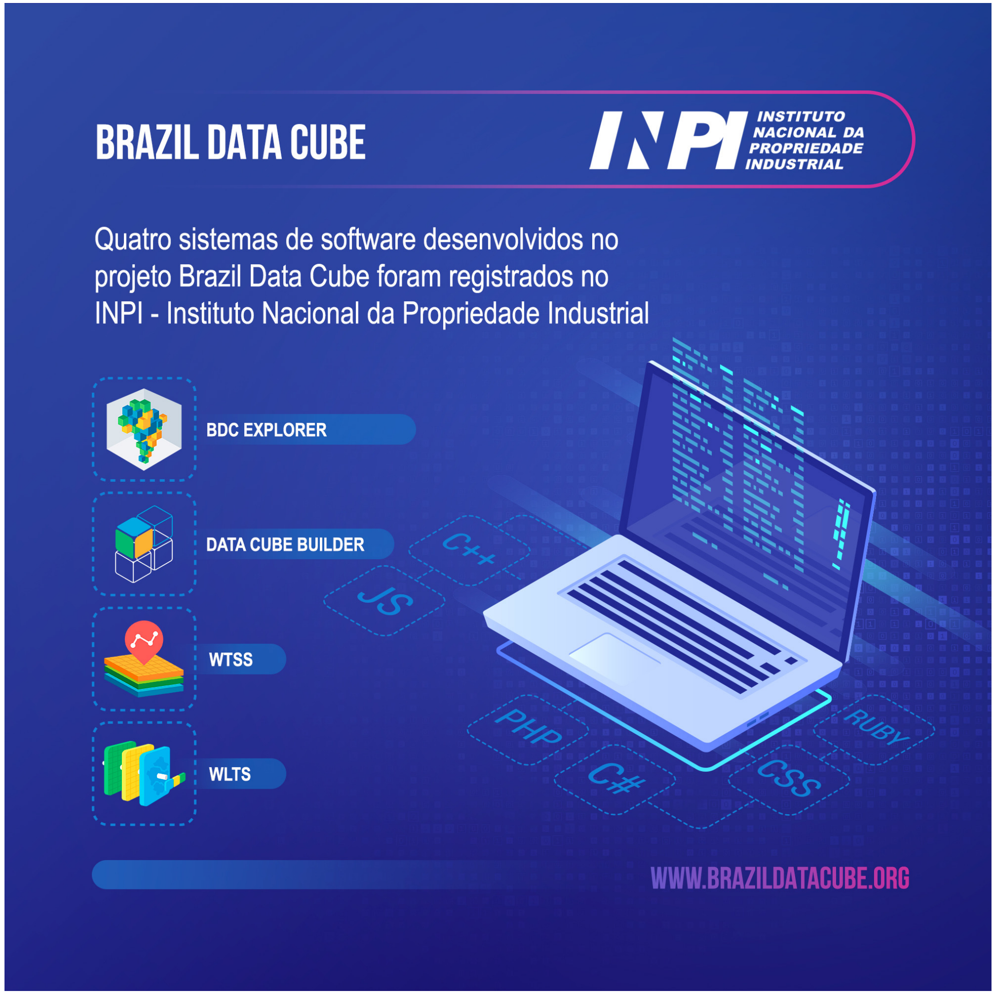
BRAZIL DATA CUBE

Computational Platform

https://github.com/brazil-data-cube

Software systems and services: 57

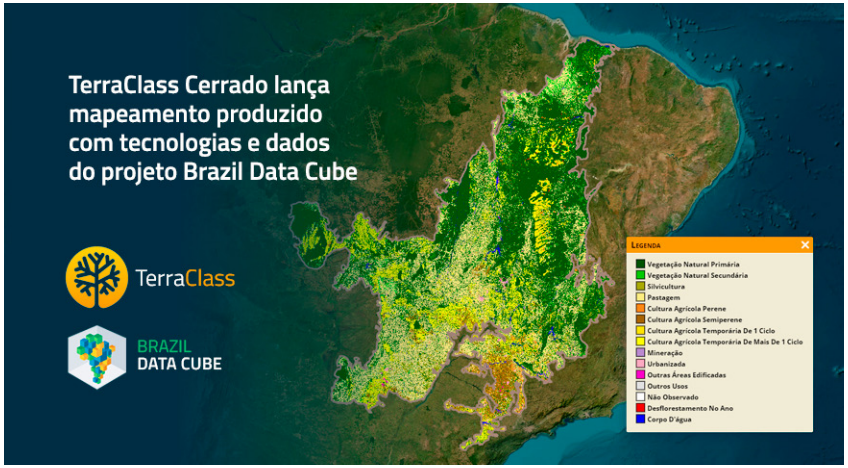
Github



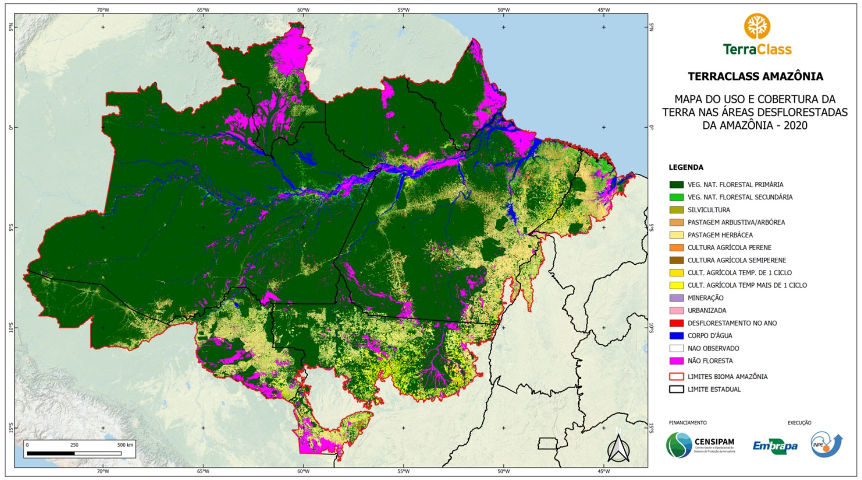
Land use and land cover maps: TerraClass project

BRAZIL
DATA CUBE

TerraClass Cerrado 2020 (Launched in December 2022)



TerraClass Amazônia 2020

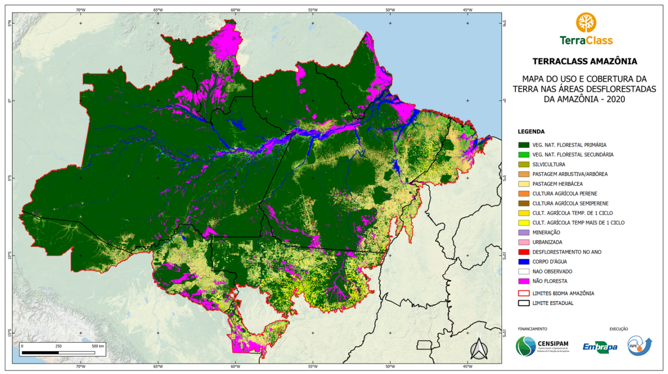


INPE
UNIDADE DE PESQUISA DO MCTI
Embrapa

Land use and land cover maps: TerraClass project

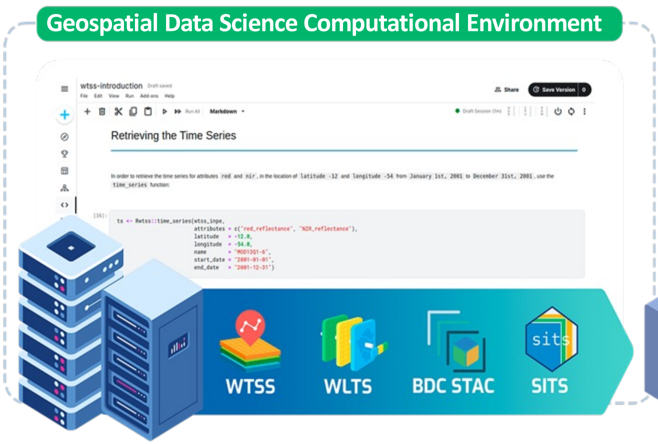
33 Terabytes (292 BDC tiles)
Sentinel-2 data cubes (16-days)
25,000 samples
Random Forest classifier

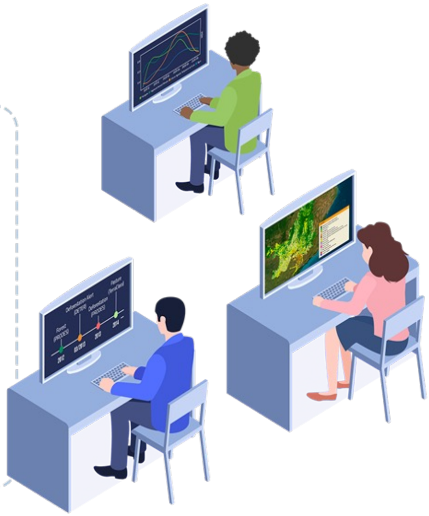
TerraClass Amazônia 2020



Land use and land cover maps: TerraClass project

BRAZIL
DATA CUBE

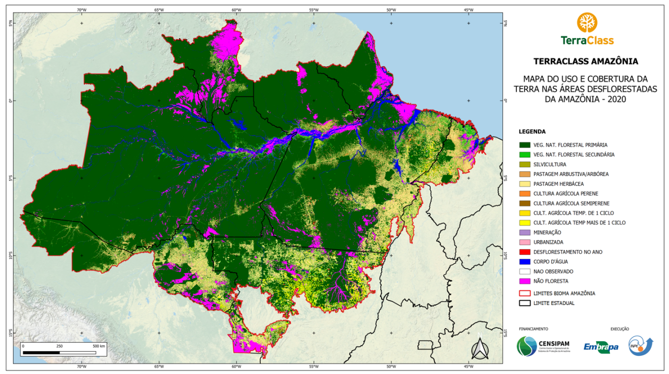




33 Terabytes (292 BDC tiles)
Sentinel-2 data cubes (16-days)
25,000 samples
Random Forest classifier

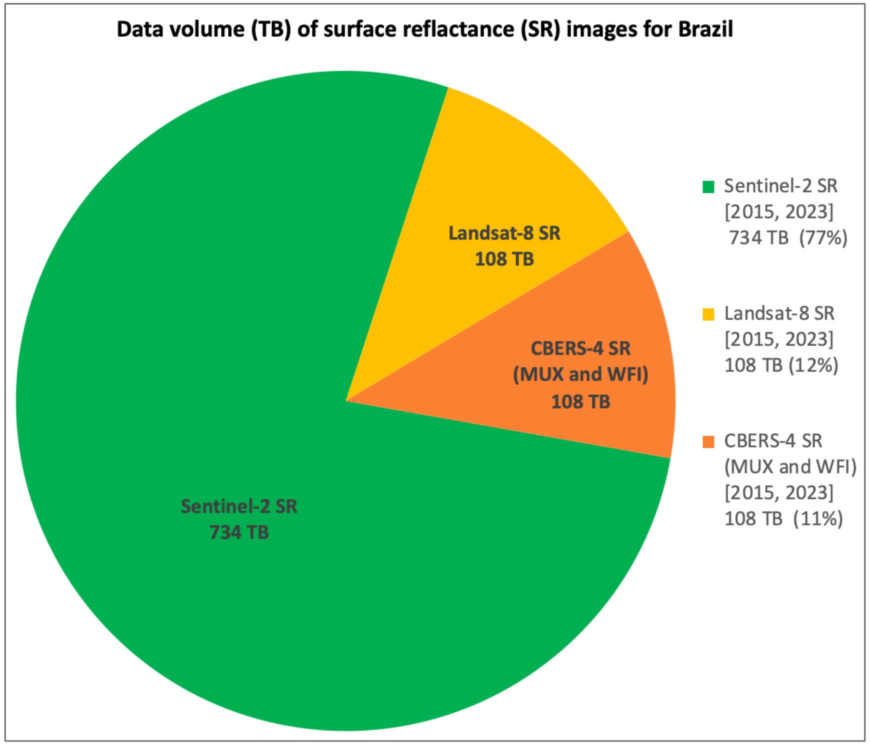
Server-side processing infrastructure at INPE

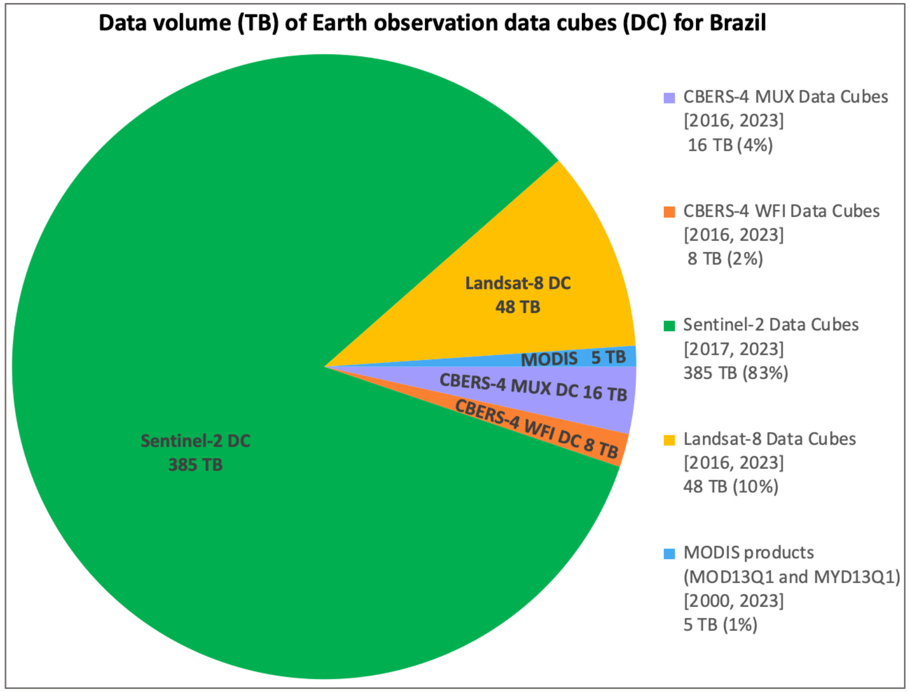
TerraClass Amazônia 2020



Challenge - Big data

~ 2 Petabytes (PB)

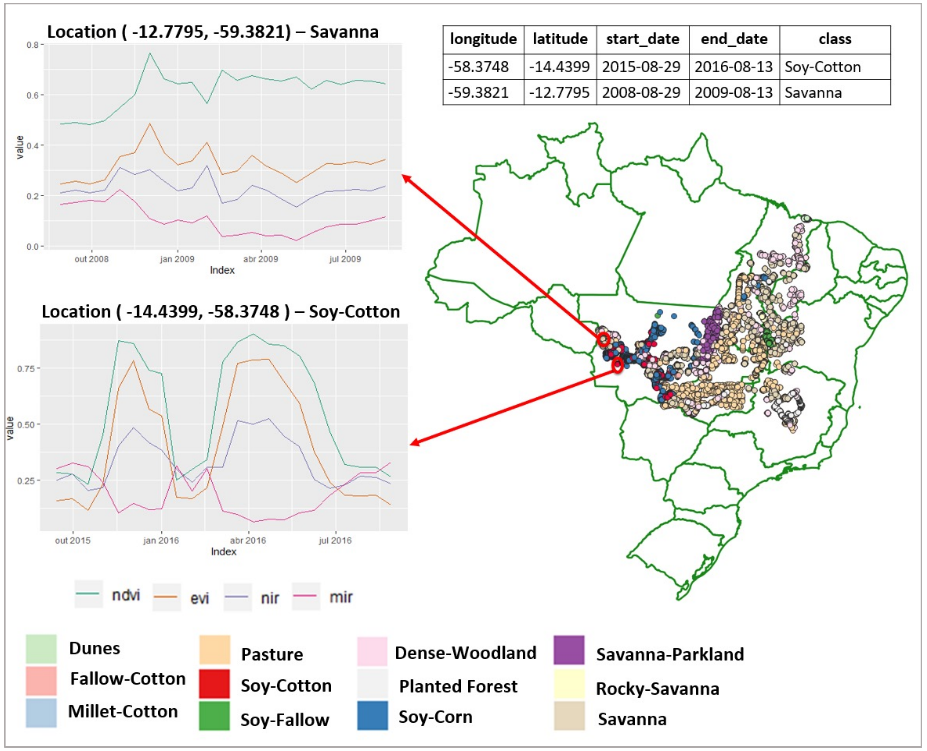


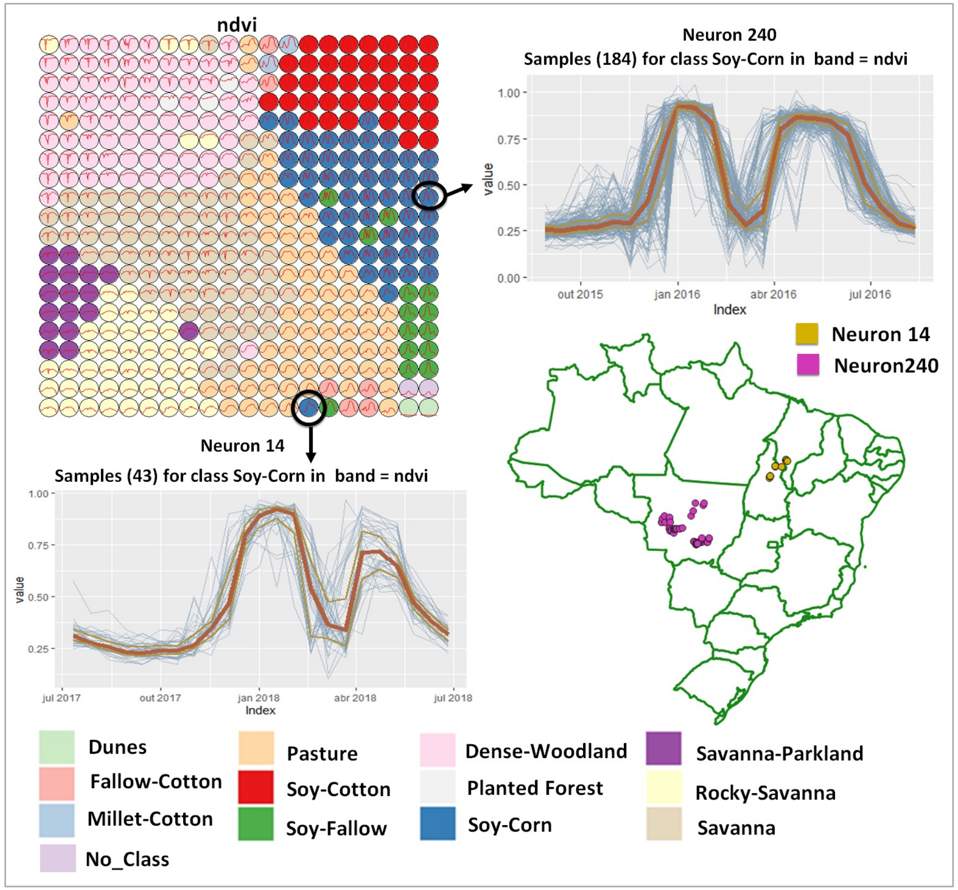


Source: [Ferreira et al., 2022]

Challenge – Samples

Methods to assess and improve the quality of land use and cover samples





Source: [Lorena Santos et al., 2021]

BRAZIL
DATA CUBE

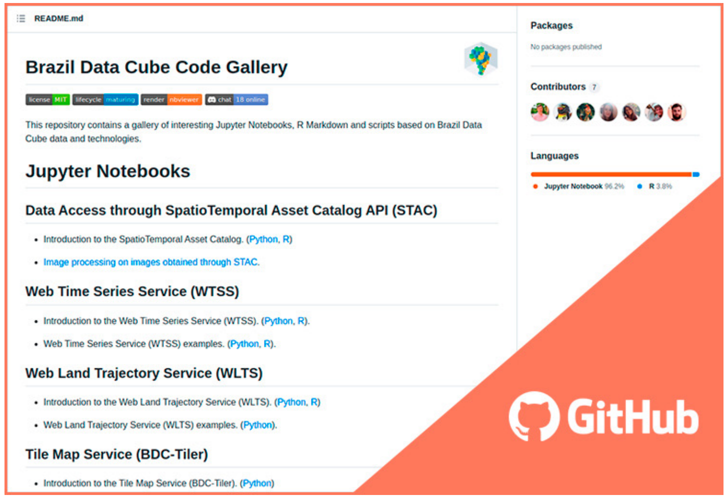
20 CAPACITAÇÕES REALIZADAS

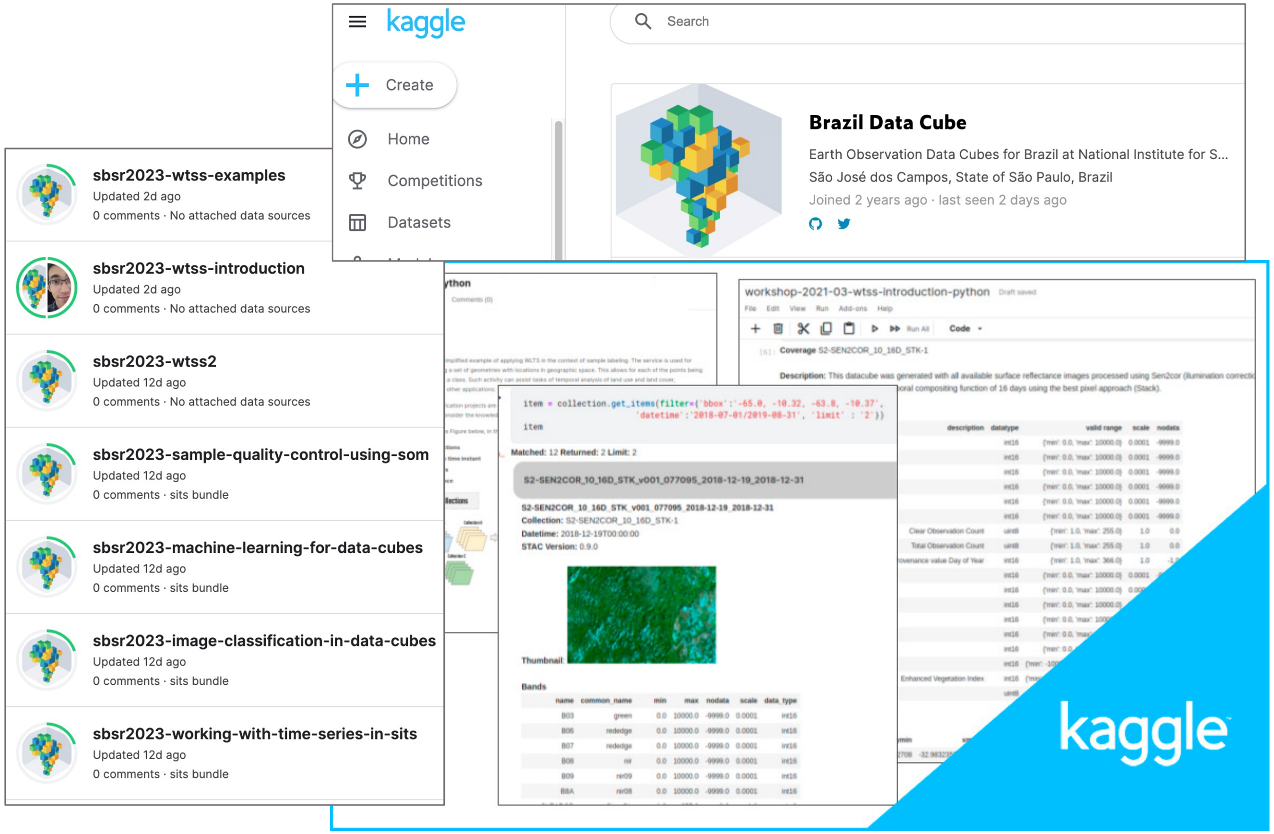
http://brazildatacube.org

≅ 1230 PARTICIPANTES

20 CAPACITAÇÕES REALIZADAS

http://brazildatacube.org

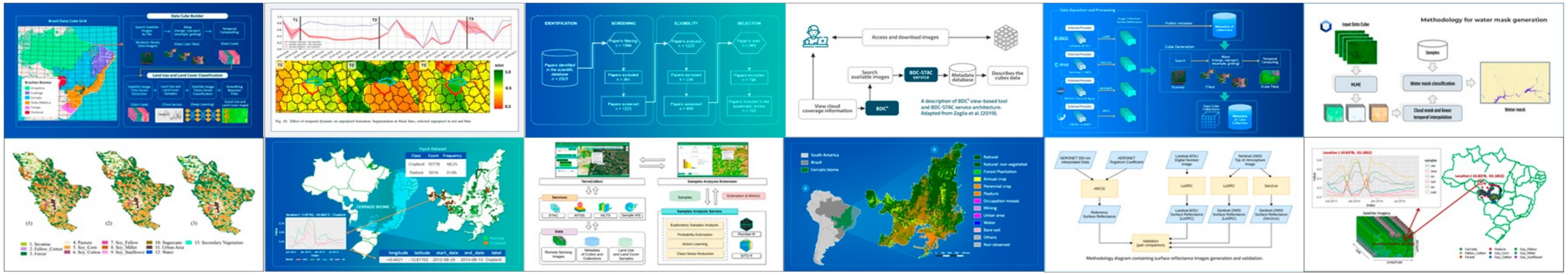




49 PUBLICAÇÕES

28 – Revistas e eventos internacionais

21 – Revistas e eventos nacionais



http://brazildatacube.org

PARTICIPAÇÃO EM 66 EVENTOS

BRAZIL DATA CUBE

Congressos, Palestras, Reunião temática, Simpósio, Workshop, Plenária, Hackathon, Cursos, Mesa redonda, Apresentações, Webinars

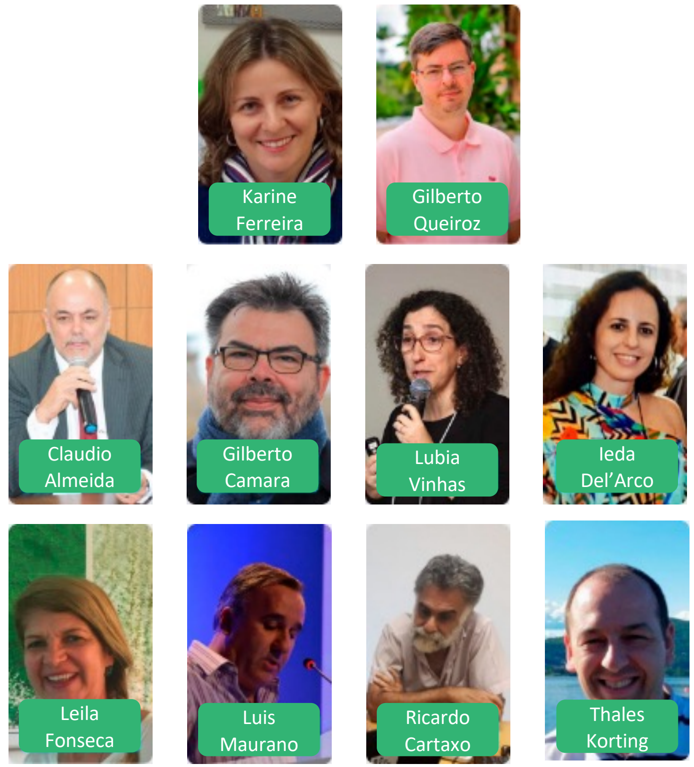
36 - Nacionais
30 - Internacionais

http://brazildatacube.org

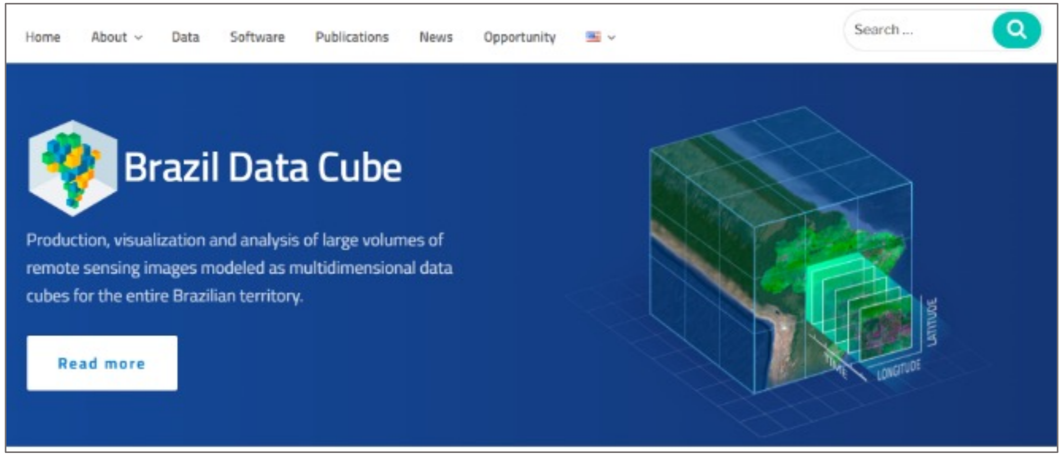
BRAZIL
DATA CUBE

Software developers, Associate researchers, Master and PhD students.

INPE Researchers



http://brazildatacube.org



TerraClass Brasil

INPE Embrapa IBGE

Projetos / Laboratórios

MAPAQUAL HARMONIZE

VISIONA Tecnologia Espacial

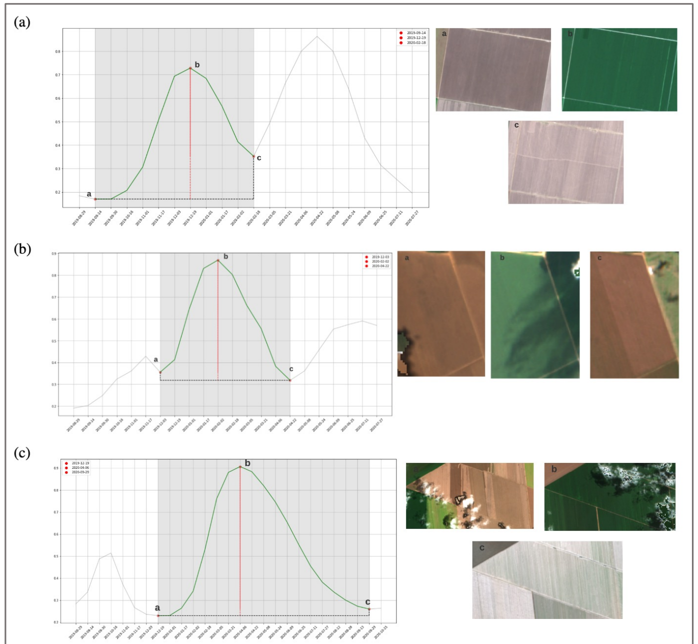
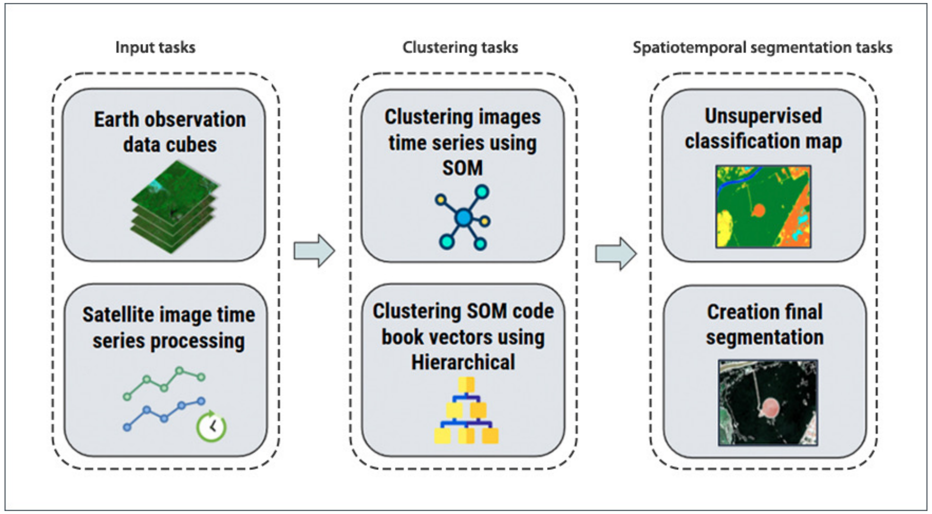


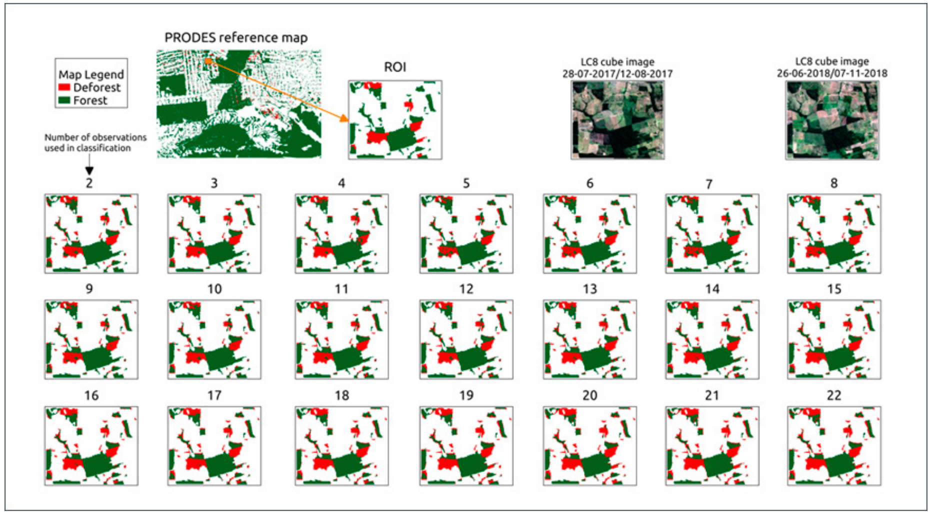
Figura 4. Métricas fenológicas de início, fim e máximo vigor vegetativo de plantio extraídas para soja (a), milho de primeira safra (b) e algodão (c) utilizando séries Sentinel-2A/B de NDVI a cada 16 dias.

BDC + CAP: Research and technological innovation



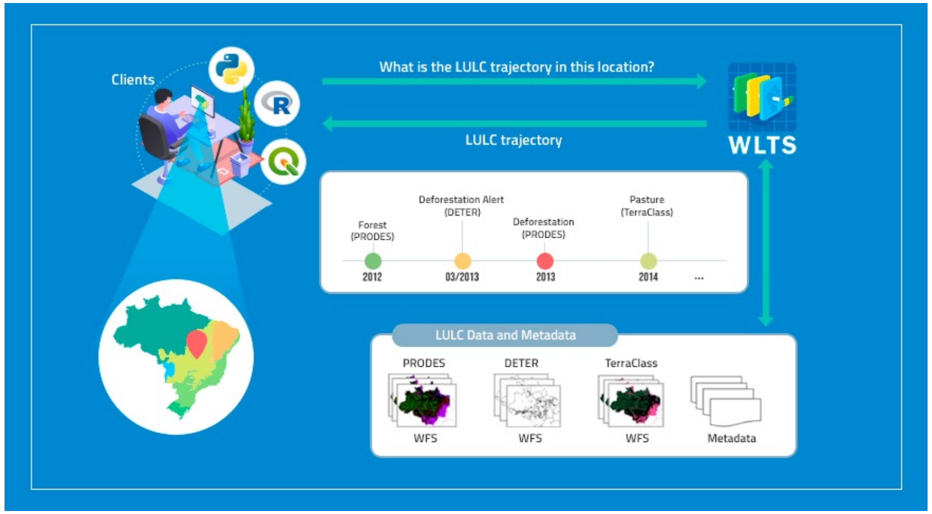
Silva, B. L. C., Souza, F. C., Ferreira, K. R., Queiroz, G. R., and Santos, L. A.: **SPATIOTEMPORAL SEGMENTATION OF SATELLITE IMAGE TIME SERIES USING SELF-ORGANIZING MAP**, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., V-3-2022, 255–261, https://doi.org/10.5194/isprs-annals-V-3-2022-255-2022, 2022.

Vieira, L. S., Queiroz, G. R., and Shiguemori, E. H.: **AN ANALYSIS OF THE INFLUENCE OF THE NUMBER OF OBSERVATIONS IN A RANDOM FOREST TIME SERIES CLASSIFICATION TO MAP THE FOREST AND DEFORESTATION IN THE BRAZILIAN AMAZON**, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B3-2022, 721–728, https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-721-2022, 2022.



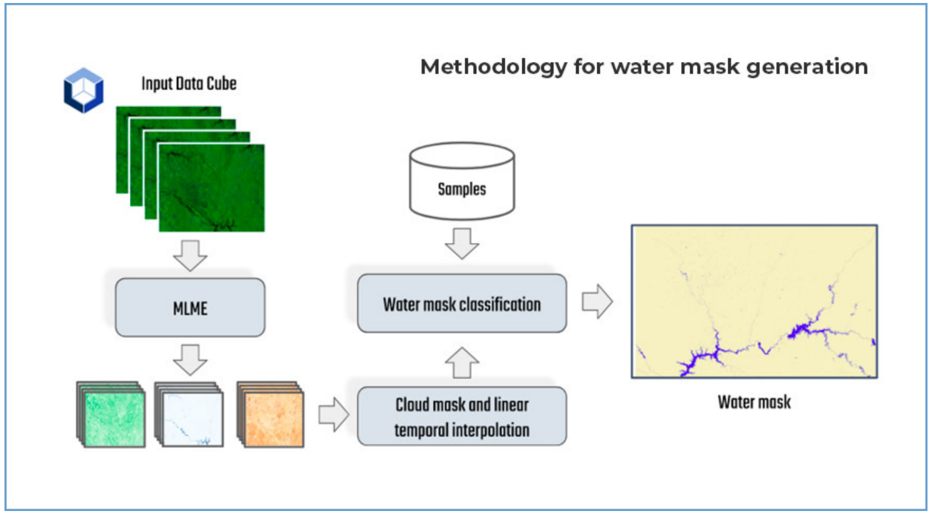
http://brazildatacube.org/en/publications/

BDC + CAP: Research and technological innovation



Zioti, F.; Ferreira, K. R.; Queiroz, G. R.; Neves, A. K.; Carlos, F. M.; Souza, F. C.; Santos, L. A.; Simoes, R. E. O. **A platform for land use and land cover data integration and trajectory analysis**. International Journal of Applied Earth Observation and Geoinformation. V 106, P 102655, Feb 2022.

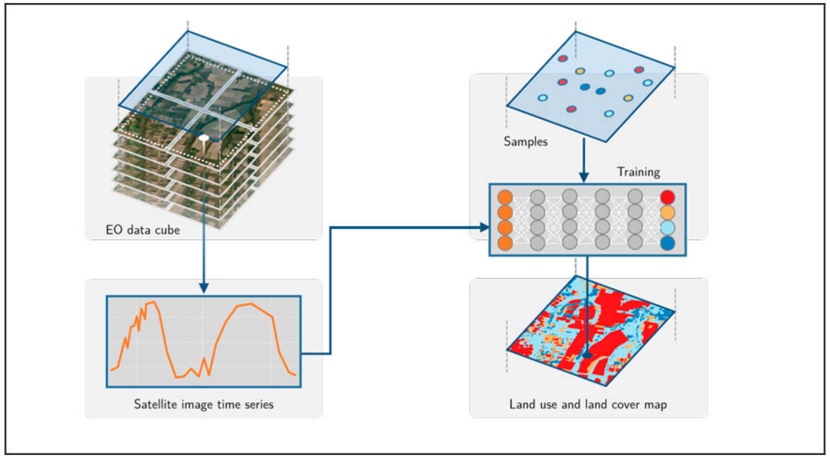
Carlos, F.M., Gomes, V.C.F., Queiroz, G.R., Souza, F.C., Ferreira, K.R., Santos, R.. **Integrating Open Data Cube and Brazil Data Cube platforms for land use and cover classifications**. Revista Brasileira de Cartografia, v73, 1036–1047, 2021

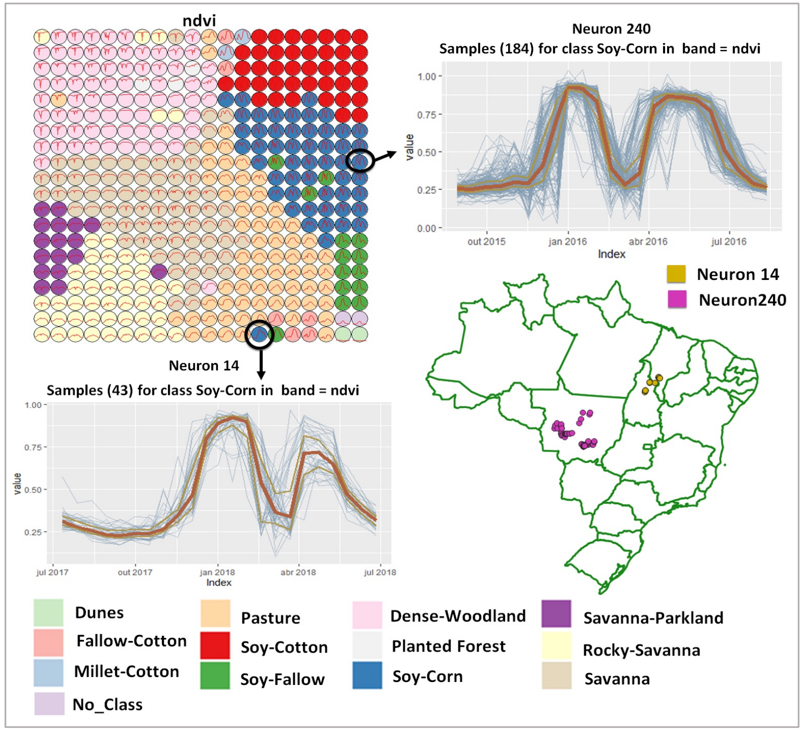


http://brazildatacube.org/en/publications/

BDC + CAP: Research and technological innovation

Santos, L.A.; Ferreira, K.; Camara, G.; Picoli, M.; Simões, Rolf. E. **Quality control and class noise reduction of satellite image time series**. ISPRS Journal of Photogrammetry and Remote Sensing. Volume 177, July 2021, Pages 75-88

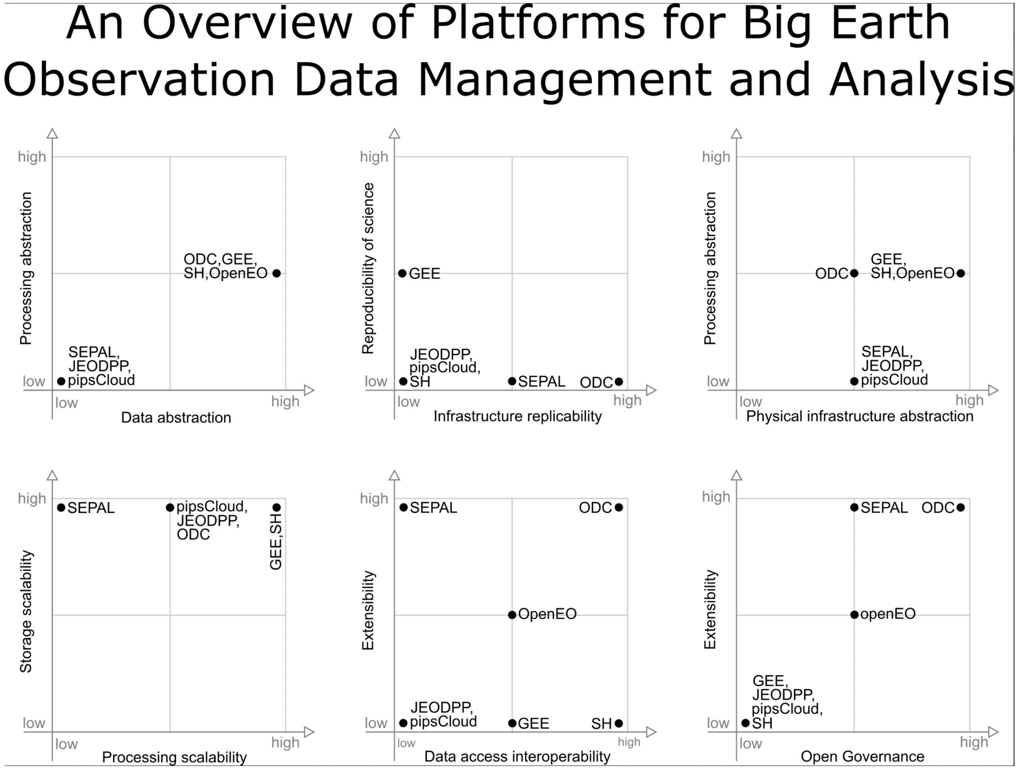




Simoes, R.; Camara, G.; Queiroz, G.; Souza, F.; Andrade, P.R.; Santos, L.; Carvalho, A.; Ferreira, K. **Satellite Image Time Series Analysis for Big Earth Observation Data**. Remote Sens. 2021, 13, 2428.

http://brazildatacube.org/en/publications/

BDC + CAP: Research and technological innovation



Gomes, V.C.F.; Queiroz, G.R.; Ferreira, K.R. **An Overview of Platforms for Big Earth Observation Data Management and Analysis**. Remote Sens. 2020, 12, 1253.

96 citations!!!! (September 13, 2022)

http://brazildatacube.org/en/publications/

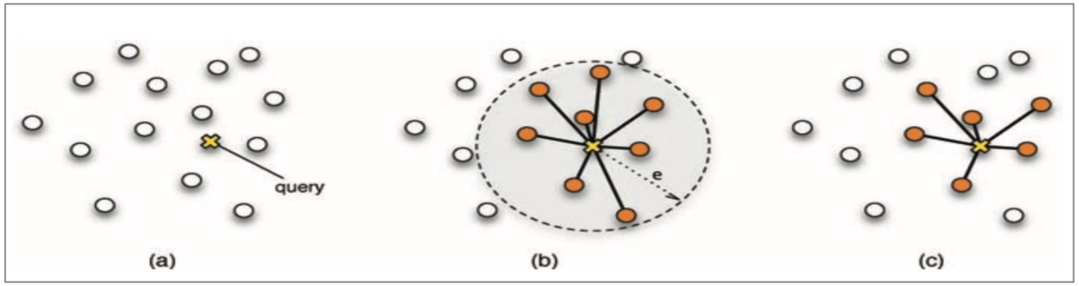
Time Series Analysis

Time Series Data Mining – Main Tasks

- Preprocessing
- Query by Content
- Clustering
- Classification
- Segmentation
- Prediction
- Anomaly Detection
- Motif Discovery

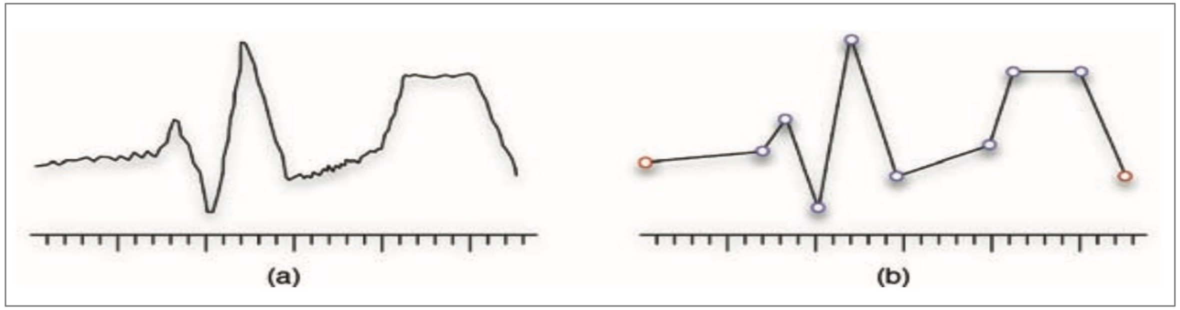
Source: (Esling and Agon, 2012)

Time Series Data Mining – Main Tasks



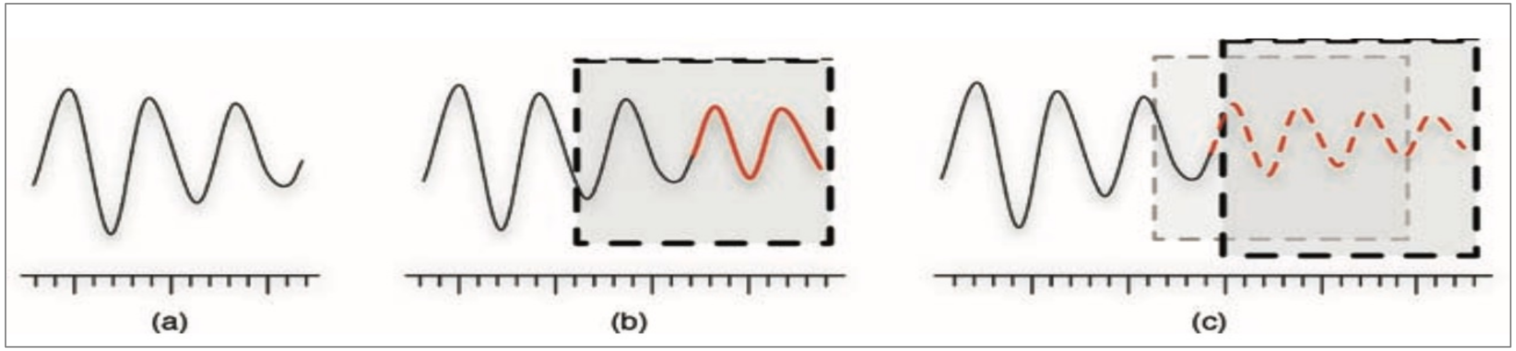
Query by content:
(a) query representation;
(b) ε-range query – distance ε
(c) K-Nearest Neighbors query.

Segmentation: the goal is to find the closest approximation of the input time series with the maximal **dimensionality reduction** factor without losing any of its essential features.



Source: (Esling and Agon, 2012)

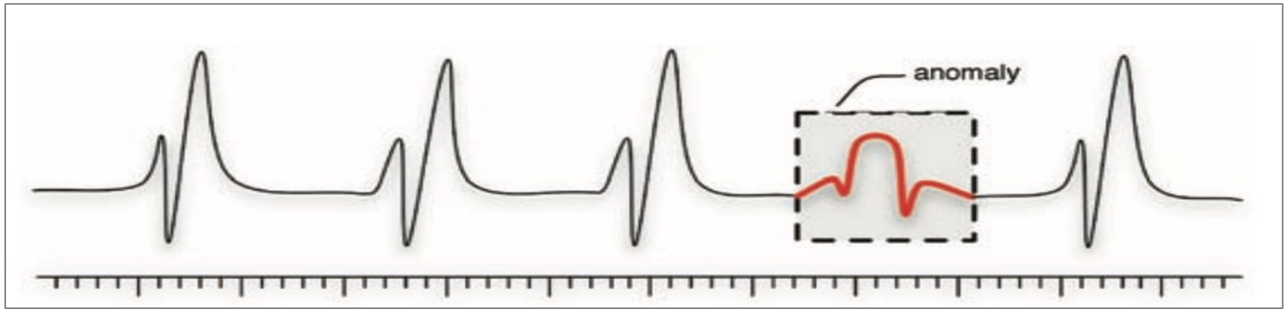
Time Series Data Mining – Main Tasks



Predition:
(a) The input time series may exhibit a periodical and thus predictable structure. (b) The goal is to forecast a maximum number of upcoming data points within a prediction window. (c) The task becomes really hard when it comes to having recursive prediction, that is, the long-term prediction of a time series implies reusing the earlier forecast values as inputs in order to go on predicting.

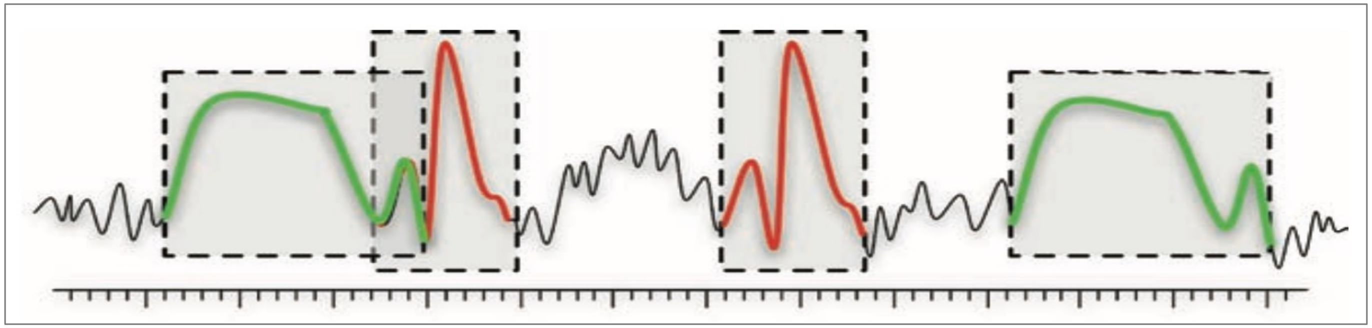
Source: (Esling and Agon, 2012)

Time Series Data Mining – Main Tasks



Anomaly Detection: a long time series which exhibits some kind of periodical structure can be modeled thanks to a reduced pattern of "standard" behavior. The goal is thus to find subsequences that do not follow the model and may therefore be considered as anomalies

Motif Discovery: consists in finding every subsequence that appears recurrently in a longer time series. These subsequences are named ***motifs***. This task exhibits a high combinatorial complexity as several motifs can exist within a single series, motifs can be of various lengths, and even overlap.



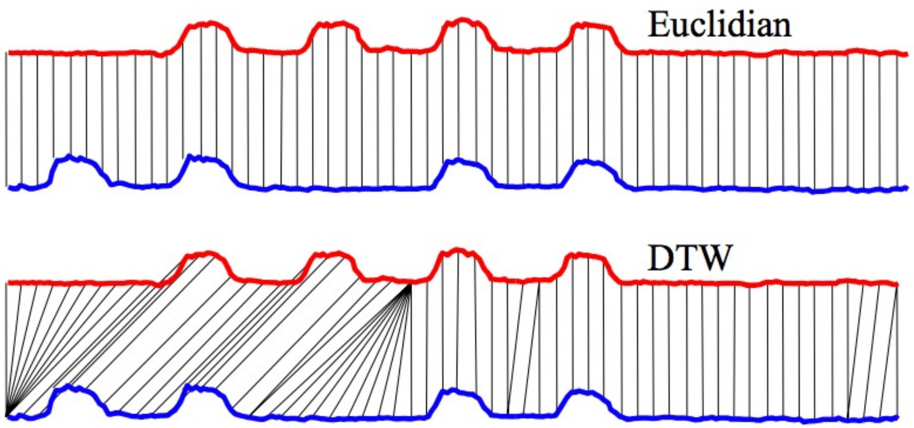
Source: (Esling and Agon, 2012)

Satellite Image Time Series Analysis – Similarity Measures

Distance Measure	Characteristics
Euclidean Distace (ED)	Lock-step Measure (one-to-one) using in indexing, clustering and classification, Sensitive to scaling.
Dynamic Time Warping (DTW)	Elastic Measure (one-to-many/one-to-none) Very well in deal with temporal drift. Better accuracy than Euclidean distance. Lowe efficiency than Euclidean distance and triangle similarity.
Longest Common Sub-Sequence (LCSS)	Noise robustness
Minimal Variance Matching (MVM)	Automatically skips outliers
Edit Distance on Real sequence (EDR)	Elastic measure (one-to-many/one-to-none), uses a threshold pattern
Cross-correlation based distances	Noise reduction, able to summarize the temporal structure
Edit Distance with Real Penalty (ERP)	Robust to noise, shifts and scaling of data, a constant reference point is used
Histogram-based	Using multi-scale time-series histograms
DISSIM	Proper for different sampling rates
Sequence Weighted Alignment model (Swale)	Similarity score based on both match rewards and mismatch penalties.
Triangle similarity measure	Can deal with noise, amplitude scaling very well and deal with offset translation, linear drift well in some situations.

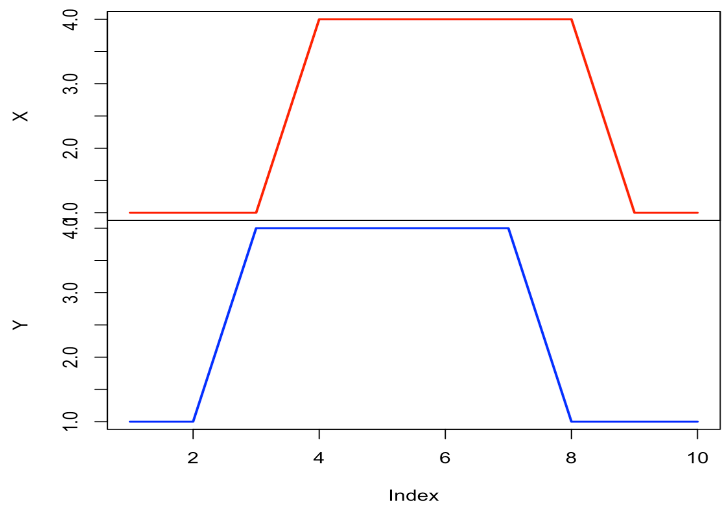
Source: (Aghabozorgi et al. 2015)

Satellite Image Time Series Analysis – DTW x Euclidean



Source: (Aghabozorgi et al. 2015)

The choice of a proper distance approach depends on the **objective**!




```r
# Two time series
X <- c(1,1,1,4,4,4,4,4,1,1)
Y <- c(1,1,4,4,4,4,4,1,1,1)

# Euclidean distance
TSdist::EuclideanDistance(X, Y)
```

```
[1] 4.242641
```

```r
# Two time series
X <- c(1,1,1,4,4,4,4,4,1,1)
Y <- c(1,1,4,4,4,4,4,1,1,1)

# Euclidean distance
TSdist::DTWDistance(X, Y)
```

```
[1] 0
```

BRAZIL DATA CUBE
brazildatacube.org

brazildatacube.org

brazildatacube.dpi.inpe.br/portal/explore

Karine R. Ferreira

OBRIGADA!

@brazildatacube

FUNDO AMAZÔNIA

FUNCATE
Fundação de Ciência, Aplicações e Tecnologia Espaciais

BNDES

INPE

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES