

CBERS DATA CUBE: A POWERFUL TECHNOLOGY FOR MAPPING AND MONITORING BRAZILIAN BIOMES

M. C. A. Picoli¹*, R. Simoes¹, M. Chaves¹, L. A. Santos¹, A. Sanchez¹, A. Soares¹, I. D. Sanches¹,
K. R. Ferreira¹, G. R. Queiroz¹

¹ General Coordination of Earth Observation (OBT) - Brazil's National Institute for Space Research (INPE)
Av. dos Astronautas, 1758, São José dos Campos - SP, Brazil
(michelle.picoli, rolf.simoes, michel.chaves, lorena.santos, alber.ipia, anderson.soares, ieda.sanches,
karine.ferreira, gilberto.queiroz)@inpe.br

KEY WORDS: Analysis Ready Data, Earth observations, information extraction, LULC classification, time series, Random Forest.

ABSTRACT:

Currently, the overwhelming amount of Earth Observation data demands new solutions regarding processing and storage. To reduce the amount of time spent in searching, downloading and pre-processing data, the remote Sensing community is coming to an agreement on the minimum amount of corrections satellite images must convey in order to reach the broadest range of applications. Satellite imagery meeting such criteria (which usually include atmospheric, radiometric and topographic corrections) are generically called Analysis Ready Data (ARD). Furthermore, ARD is being assembled into multidimensional data cubes, minimising pre-processing tasks and allowing scientists and users in general to focus on analysis. A particular instance of this is the Brazil Data Cube (BDC) project, which is processing remote sensing images of medium spatial resolution into ARD datasets and assembling them as multidimensional cubes of the Brazilian territory. For example, BDC users are released from performing tasks such as image co-registration, aerosol interference correction. This work presents a BDC proof of concept, by analysing a BDC data cube made with images from the fourth China-Brazil Earth Resources Satellite (CBERS-4) of one of the largest biodiversity hotspot in the world, the *Cerrado* biome. It also shows how to map and monitor land use and land cover using the CBERS data cube. We demonstrate that the CBERS data cube is effective in resolving land use and land cover issues to meet local and national needs related to the landscape dynamics, including deforestation, carbon emissions, and public policies.

1. INTRODUCTION

Currently, new satellite and terrestrial remote sensing systems produce such large amounts of data at fine temporal and radiometric resolutions (Nativi et al., 2015) that new data storing solutions are dearly needed. Besides, the time required to pre-process such amount of data diminishes the time actually invested in analysis. Analysis Ready Data (ARD) and multi-dimensional data cubes are preconditions to fulfil new analysis demands and to increase the levels of detail and accuracy required in the a fast changing environment (Nativi et al., 2017).

The purpose of Earth observation (EO) data cubes is to organise the data to make their use so simply and intuitive that users can focus on developing and testing their methods (Appel, Pebesma, 2019). An EO data cube is a four-dimensional array relating dimensions to longitude, latitude, time, and spectral bands (Appel, Pebesma, 2019). Around the world, EO data cubes initiatives are providing open data for the common good of society (Killough, 2018).

The first EO data cube was the Australian Geoscience Data Cube (AGDC) (Dhu et al., 2017, Lewis et al., 2017), which still produces and distributes Analysis Ready Data made from Landsat images, enabling users to explore and increase the impact of EO data (Committee on Earth Observation Sciences (CEOS), n.d.). The AGDC runs on top of the Open Data Cube (ODC) infrastructure, which is been used to create other national and regional data cubes such as those of Switzerland (Giuliani et al., 2017), Colombia (Ariza-Porrás et al., 2017, Bravo

et al., 2017), Africa (Killough, 2019), Armenia (Asmaryan et al., 2019), China (Yao et al., 2018), and Mexico (Dhu et al., 2019).

Other initiatives produce data cubes by combining images from different sensors, such as the Cubesat Enabled Spatio-Temporal Enhancement Method (CESTEM) and the Framework for Operational Radiometric Correction for Environmental monitoring (FORCE). CESTEM produces radiometric harmonisation of images from Planet's constellation, Landsat, Sentinel, and MODIS (Houborg, McCabe, 2018) while FORCE harmonises ARD data from Landsat 8 Operational Land Imager (OLI) and Sentinel-2 MultiSpectral Instrument (MSI) using a software developed by Frantz (2019).

The Brazil National Institute for Space Research (INPE) is developing the Brazil Data Cube (BDC) project to create ARD data cubes of the Brazilian territory¹. Among others data cubes, BDC provides one made of images from the fourth China-Brazil Earth Resources Satellite (CBERS-4). CBERS-4 sensors provide medium resolution images in the visible and infrared region of the electromagnetic spectrum.

A medium resolution CBERS-4 data cube could extend and improve the scope of environmental monitoring studies such as deforestation mapping, greenhouse gas emission assessment, and forest fire detection. In addition, it may encourage the scientific community to research and develop new cartographic products since several works have established data cubes as a technology for mapping land use and land cover (LULC) (Hamunyela

* Corresponding author

¹ For more details, see <http://brazildatacube.org/>

et al., 2016, Brooke et al., 2017, Lucas et al., 2019, Xi et al., 2019) and for monitoring both forest (Hamunyela et al., 2017, Hermosilla et al., 2018) and urbanisation (Killough, 2019).

In this study, we demonstrate the applicability of the CBERS-4 data cube for mapping LULC. Additionally, we present to the scientific community a data cube that is already available on the BDC project platform.

2. DATA CUBE INITIATIVES

The Australian Geoscience Data Cube (AGDC) has been producing ARD data from the Landsat and Sentinel-2/MSI satellites since 2018. Its aim is to explore the EO potential by addressing the volume, velocity, and data diversity challenges of EO big data (Lewis et al., 2017). The AGDC applications range from understanding environmental changes, such as water availability and urban and agricultural expansion, to allow companies and industries to have access to EO data so that they can innovate and develop new products (Dhu et al., 2019).

The purpose of the Swiss data cube (SDC) is to monitor environmental changes, spatially and temporally, and to provide ARD data so that Swiss scientific institutions can innovate and generate information that improves knowledge about the Swiss environment, in addition to enabling more effective responses to problems of national importance (Giuliani et al., 2017). The Swiss Data Cube (SDC) contains optical data from the Landsat and Sentinel-2 satellites and radar data from the Sentinel-1 (Dhu et al., 2019).

The Regional Data Cube for Africa (ARDC), launched in 2018, comprises five countries: Ghana, Kenya, Senegal, Sierra Leone, and Tanzania. ARDC's goal is to provide access to EO data, in free and open data infrastructure, to address the United Nations Sustainable Development Goals (SDG), and other priorities in each country (Killough, 2019). In addition, to capacity end-users to apply EO ARD to address local and national needs. The ARDC includes Landsat ARD, since 2000, but plan to add Sentinel-1 and Sentinel-2/MSI data (Dhu et al., 2019).

The objective of the Brazil Data Cube (BDC) project, started in 2019, is to develop a platform for the analysis and visualisation of large volumes of EO ARD and harmonise data from different satellites and sensors, for the Brazilian biomes. Besides, to create LULC maps and provide support for other Brazilian projects to monitor deforestation, burning, and land use and land cover changes. The BDC consists of medium spatial resolution sensors (20-30 m) from the Landsat 8/OLI, CBERS-4/WFI, and Sentinel-2/MSI platforms, covering the entire Brazilian territory.

The Colombian Geoscience Data Cube (CDCol) aims to cover the entire life cycle of the image analysis process, and therefore provide ARD data to support Colombian institutions, which will benefit from the information to support the forest and carbon monitoring system (Ariza-Porras et al., 2017, Bravo et al., 2017). The CDCol initial input includes 15 years of satellite images of Landsat 5, 7, and 8.

Armenia developed and implemented the first version of an Armenian data cube, in partnership with the Swiss data cube, in order to obtain data that support the challenges that the country faces related to environmental issues and the lack of data (Asmaryan et al., 2019). The Armenian data cube includes Landsat

5, 7, 8, and Sentinel-2/MSI ARD over Armenia, from 2016 to 2019.

The China Data Cube (CDC) is being developed to meet the needs of researchers in related areas, such as monitoring changes in the ecosystem, floods, agriculture, climate, etc (Yao et al., 2018). The China Data Cube (CDC), which uses the Open Data Cube infrastructure, has inserted the China's GF1 satellite data and plans to include more China EO data in the CDC, such as HJ1A/1B, ZY and other satellites as Landsat.

The Mexico Geospatial Data Cube (MGDC) is being developed, in collaboration with Geoscience Australia, at the National Institute of Statistics and Geography of Mexico (INEGI). The MGDC will contain ARD Landsat images since 1984, but its architecture is prepared to receive Sentinel-2/MSI data (Dhu et al., 2019). MGDC product images have already been used to provide information on issues related to Natural Resources and Agricultural Statistics in Mexico (Dhu et al., 2019). The MGDC system is expected to be an INEGI's transversal service platform.

3. CBERS-4 DATA CUBE

The CBERS-4 data cube is one of the ARD cubes that the Brazil Data Cube project has been developing. The satellite CBERS-4 was launched on December 2014 with four sensors, Panchromatic and Multispectral camera (PAN), Multispectral Camera Regular (MUX), Wide Field Imaging Camera (WFI) and the Multispectral and Thermal Imager (IRS). Their characteristics are described in Table 1.

Sensor	Spectral bands	Resolution	Width range imaged	Revisit
PAN	0.51 – 0.85 μ m (Pan)	5m/10m	60km	5 days
	0.52 – 0.59 μ m (G)			
	0.63 – 0.69 μ m (R)			
	0.77 – 0.89 μ m (NIR)			
MUX	0.45 – 0.52 μ m (B)	20m	120km	26 days
	0.52 – 0.59 μ m (G)			
	0.63 – 0.69 μ m (R)			
	0.77 – 0.89 μ m (NIR)			
IRS	0.50 – 0.90 μ m (Pan)	40m/80m	120km	26 days
	1.55 – 1.75 μ m (SWIR)			
	2.08 – 2.35 μ m (SWIR)			
	10.40 – 12.50 μ m (TH)			
WFI	0.45 – 0.52 μ m (B)	64m	866km	5 days
	0.52 – 0.59 μ m (G)			
	0.63 – 0.69 μ m (R)			
	0.77 – 0.89 μ m (NIR)			

Table 1. Summary of the characteristics of the CBERS-4 sensors. Source: <http://www.cbbers.inpe.br/sobre/cameras/cbers3-4.php>

The CBERS-4 data cube was created using images from the WFI sensor. The WFI images have 64m of spatial resolution and the raw data are processed to produce co-registered, top of atmosphere and surface reflectance ARD images (Dwyer et al., 2018). Also, the indices NDVI and EVI, and a cloud mask product are computed (Figure 1). From surface reflectance product, we generate composite images using different time periods, one month and 16 days. These periods encompass, on average, 6 and 3 observed images, respectively. For each period, after removing cloud masked pixels, we choose

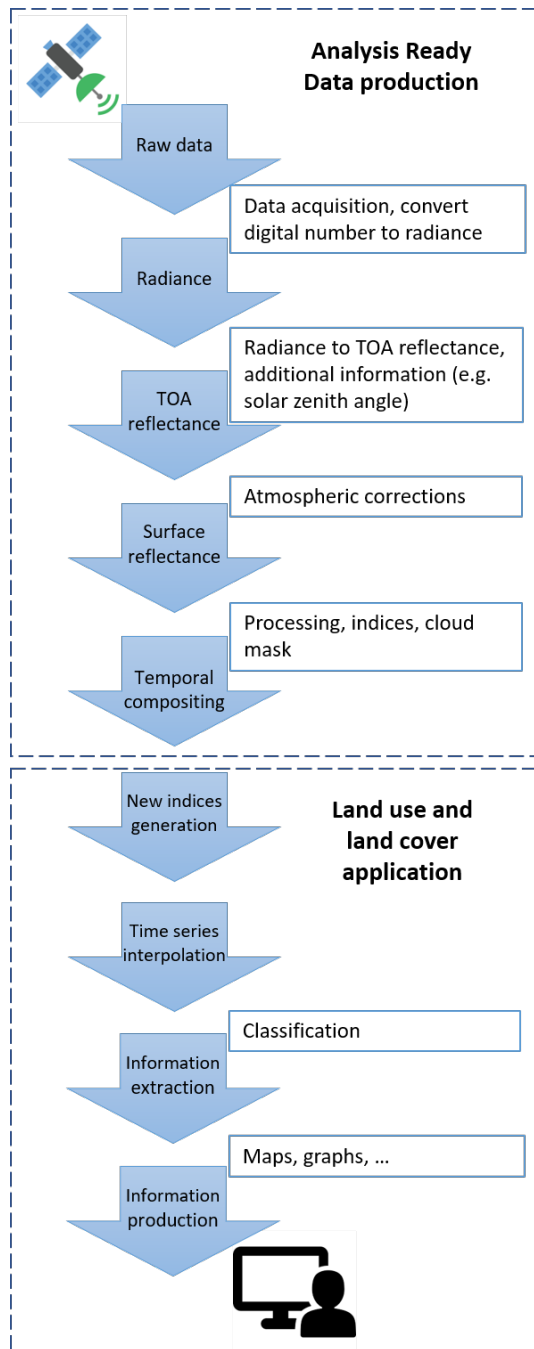


Figure 1. Data cube generation and LULC classification workflow. Above: Analysis Ready Data (ARD) production in five steps (data acquisition, conversion to radiance, top of atmosphere (TOA) reflectance, surface reflectance, and temporal compositing) provided by CBERS-4 data cube. Below: Data preparation for LULC using time series data. Source: Adapted from Giuliani et al. (2017).

one value using two different strategies, the simple median and the “stack” algorithm, that prioritises the non-cloud values of images with less cloud cover. This process generated four temporal compositing products. In this study, we used images from the monthly compositing product using “stack” algorithm.

All images are available on the BDC project portal. BDC ARD data is supplied as tiled products defined in an equal area projection. The spatial footprint comprehends 1 : 250,000 grid with

a tile size of approximately $165 \times 110 km$. All data is provided in Cloud Optimised GeoTIFF (COG) format (Cloud Optimized GeoTIFF, 2019) and described according to the Spatiotemporal Asset Catalog (STAC) (Spatio Temporal Asset Catalog, 2019) specification. The software used to build, access, and process the CBERS-4 data cubes is open source and the code is available on the BDC project’s Github at <https://github.com/brazil-data-cube>.

4. PROOF OF CONCEPT

In this section, we present an application of LULC classification. After delimiting the study area, the process consisted of preparing the data cubes, obtaining the time series for the samples, training and validating the random forest model, and generating the LULC map (Figure 2).

4.1 Study area

The study area is located in the *Cerrado* biome, on the border of three Brazilian states: *Mato Grosso*, *Mato Grosso do Sul* and *Goiás*, Brazil (Figure 3). The period of analysis is from September of 2018 to August of 2019.

The *Cerrado* is the second largest biome in South America, occupying an area of 2 million km^2 , about 22% of the Brazilian territory. Considered as a global hotspot of biodiversity (Klink, Machado, 2005), it presents an extreme abundance of endemic species but it suffers an exceptional loss of habitat. Because of its biological diversity, the *Cerrado* is recognised as the richest savanna in the world, sheltering at least 11,600 species of catalogued native plants (Brazilian Ministry of Environment, 2019).

4.2 Data cube access and preparation

The CBERS-4 data cube can be accessed by the BDC portal or by STAC service, which provides information about the images

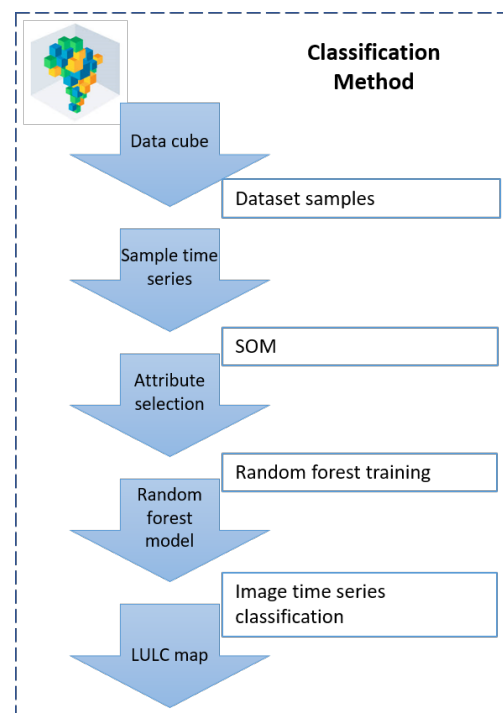


Figure 2. Land use and land cover classification method.

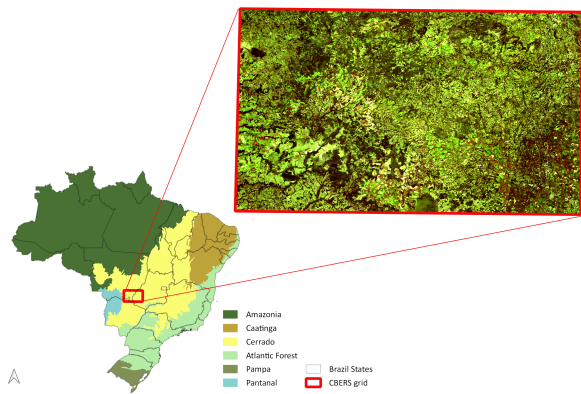


Figure 3. Left: study area in relation to Brazil and its biomes. Right: CBERS-4/WFI image from June, 2018 of the study area (RGB - 14/13/15 bands).

as well as the links to access their contents. The service API enables automated access and preparation of the data cube. Here, we computed new indices and filled null data using temporal linear interpolation. We used the monthly composed images data cube using the “stack” algorithm.

In addition to the two indices provided by this product (Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI)), we computed other four indices: Global Environment Monitoring Index (GEMI), Green Normalized Difference Vegetation Index (GNDVI), Normalized Difference Water Index (NDWI2), Photosynthetic Vigour Ratio (PVR). Resulting in six vegetation indices and four spectral bands: blue, green, red, and near-infrared bands (Table 2).

Index	Equation	Source
EVI	$\frac{2.5(nir-red)}{(nir+6red-7.5blue)+1}$	(Huete et al., 1997)
GEMI	$\eta(1-0.25\eta) - \frac{(red-0.125)}{(1-red)}$	
	$\eta = \frac{(2(nir^2-red^2)+1.5 \times nir+0.5 \times red)}{(nir+red+0.5)}$	(Pinty, Verstraete, 1992)
GNDVI	$\frac{(nir-green)}{(nir+green)}$	(Gitelson et al., 1996)
NDVI	$\frac{(nir-red)}{(nir+red)}$	(Rouse et al., 1974)
NDWI2	$\frac{(green-nir)}{(green+nir)}$	(McFeeters, 1996)
PVR	$\frac{(green-red)}{(green+red)}$	(Metternicht, 2003)

Table 2. Summary of the vegetation indices.

The CBERS-4 data cube consists of images organised by date and band. Organised by date, the images are stacked forming time series for each attribute of each pixel. To fill eventual null data in the time series, we interpolated the lacking data linearly using the closest valid values available. After this processing, we extracted the time series for each sample for the training stage and used it later to generate the LULC maps.

4.3 Sample dataset

We merged two sample dataset, one coming from high resolution images, collected by Remote Sensing specialist, and the

other from the BDC project database. This dataset includes 1,042 LULC samples divided into four classes: Natural Vegetation (NV), Pasture (P), Semi-Perennial Crop (SP Crop), and Annual and Perennial Crop (AP Crop).

To evaluate the separability of these classes in the sample dataset, we used a neural network Self-Organizing Map (SOM). SOM is a suitable clustering method when working with time-series data (Aghabozorgi et al., 2015), it is an unsupervised neural network where the input layer is the sample dataset, and the output layer is a data set of grouped neurons.

SOM allows us to map from high to low dimension spaces while preserving the topology of the data and reducing the computational cost. SOM evaluates which spectral bands and vegetation indexes are best suited for LULC separability (Santos et al., 2020). The SOM parameters were: a grid size of 5×6 , a learning rate decreasing from 0.05 to 0.01, and 300 iterations.

4.4 Random forest classification

The random forest algorithm (Breiman, 2001) is an ensemble method based on a decision tree model. Its strategy consists in developing many decision trees via bootstrap and random feature selection to reduce classification bias. The majority voting schema is used to obtain the final classification model.

The classification was done using the *sits* R package (Camara et al., 2018), an open source software developed by our research group. For the classification, we used the full depth of CBERS image time series to create larger dimensional spaces, and we set the number of decision trees to 1000. At each growing tree, only a fraction of the attributes are considered to split a node according to the *Gini* index, used here as an attribute relevance criterion.

To validate the resulting classification we used a 5-fold cross-validation (Wiens et al., 2008). It ran five different assessments using 80% of the samples for training and 20% for prediction. The average accuracy of the five classifications was used to produce a single accuracy estimation.

5. RESULTS AND DISCUSSION

For the study area samples, the SOM clustering reduced the size of the sampling dataset by 17.8%, going from 1,042 to 856 samples. We used this filtered dataset to train the classification model. The user’s and producer’s accuracy for the LULC classes mapped are presented in Table 3. The classification quality assessment using 5-fold cross-validation (Wiens et al., 2008) of the training samples showed an overall accuracy of 97.0% and a Kappa index of 0.96.

The results demonstrate that the CBERS-4/WFI data cube is not only ideal for mapping LULC but also for detecting land use and land cover change (LULCC), as it provides ARD image time series. Currently, one of the biggest worries of the *Cerrado* biome is the rapid change occurred in LULC (Soterroni et al., 2019). Agricultural crops and pasture areas have been expanding over natural vegetation with considerable speed (Picoli et al., 2020). Maps produced from data cube in the ARD format can be a powerful tool for monitoring the dynamics of land use and land cover in the *Cerrado*.

Besides, the CBERS-4/WFI data cube can be considered better than MODIS products - widely used for mapping the *Cerrado* biome - because of its spatial resolution. For example,

	AP Crop	NV	P	SP Crop	UA
AP Crop	250	0	1	6	0.97
NV	0	245	0	0	1.00
P	0	0	211	5	0.98
SP Crop	6	0	5	123	0.92
PA	0.98	1.00	0.97	0.92	

Table 3. Confusion matrix of CBERS-4 data cube obtained by 5-fold cross validation of classification of field data, and values of producer's accuracy (PA) and user's accuracy (UA) for each class: Annual and Perennial Crop (AP Crop), Natural Vegetation (NA), Pasture (P), and Semi-Perennial Crop (SP Crop).

the spatial resolution of MODIS (250 x 250 m) causes spectral mixing and limits pattern recognition in heterogeneous landscapes (Zhong et al., 2016) as the study region. For natural formations, MODIS is unable to capture complex vegetation gradients such as some sub-types of savanna (e.g., shrublands, and grasslands) (Schwieder et al., 2016), which turns cerrado areas spectrally similar to forest areas in the function of its density (Simoes et al., 2020). For other vegetation classes, such as pasturelands and agricultural vegetation the spatial resolution of MODIS may cause spectral confusion due to its seasonal variation (Chaves et al., 2018, Picoli et al., 2018). As showed in this work, these both situations can be improved by using the CBERS-4/WFI data cube.

Figure 4 shows the resulting map with the spatial distribution of LULC classes, from September 2018 to August 2019, using the vegetation indices derived from the CBERS-4/WFI data cube EVI, NDVI, GEMI, GNDVI, NDWI2, PVR, and the spectral bands red, green, blue, and near-infrared, applying the random forest algorithm.

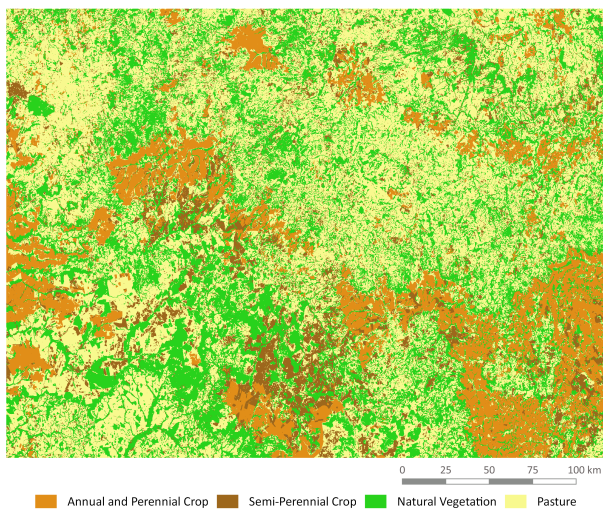


Figure 4. Random forest classification of the study area, from September 2018 to August 2019, using the CBERS-4/WFI data cube.

The CBERS-4/WFI data cube can be used to monitor the expansion of agriculture, pasture, and urban areas, natural disasters, and to map deforestation. Projects like the Monitoring of the Brazilian Amazon Deforestation by Satellite (PRODES) (Brazil's National Institute for Space Research – INPE, 2020) that already use CBERS-4 satellite data to map deforestation across the Legal Amazon, can be benefited by incorporating the

CBERS-4/WFI data cube into its classification systems. Using the CBERS-4 data cube, the technicians of this project will be able to pass up the image processing steps and just focus on detecting deforestation.

Other projects for monitoring LULC in Brazil, such as Terra-Class (Almeida et al., 2016) and MapBiomass (Azevedo et al., 2018), will also benefit from data cubes, as they will have time series data in the ARD format to map Brazilian biomes. These projects could allocate more human resources in the development of mapping. Besides the benefits already mentioned, the CBERS-4/WFI data cube can support public policies aimed to mitigate the impacts of global environmental changes. For the Brazilian context, the use of this product also can symbolise more autonomy due to the national efforts and technology implemented to develop and launch the satellite in partnership with China.

6. NEXT STEPS

Recently, on December of 2019, a new satellite of the CBERS program (CBERS-04A) was successfully launched carrying sensors with the compatible specifications of those inboard of CBERS-4. This launch ensures continuity of the image capturing as well as an increment in the frequency of revisits. The BDC project intends to implement in a single data cube images from both CBERS-4 and CBERS-04A.

In the context of the BDC Project, other data cubes with medium resolution are being generated and tested besides CBERS-4/WFI, such as Sentinel-2/MSI and Landsat 8/OLI imagery. Initially, the period of these data cubes will be from 2017 (the launch of Sentinel-2B) to 2021 for the entire Brazilian territory.

Moreover, the BDC research team is studying multi-sources image harmonisation strategies to produce spectrally harmonised data cubes. The BDC project has been testing atmospheric correction algorithms and developed procedures to produce data cubes containing harmonised surface reflectance images from the Landsat 8/OLI and Sentinel-2/MSI satellites. This will increase the data frequency and adequate observations to characterise highly dynamic LULC processes.

7. CONCLUSIONS

Because of the large volume of EO data, such as the CBERS-4 WFI satellite, which has data every 5 days since 2014, users demand a lot of time to process and organise images. Since the Brazil Data Cube project, users now have access to use EO ARD to meet their needs.

The proof of concept presented in this paper show that the LULC classification of the studied area using the EVI, NDVI, GEMI, GNDVI, NDWI2, and PVR vegetation indices, and the red, green, blue and near-infrared spectral bands is promising for LULC mapping and monitoring. The map generated from the CBERS-4 data cube provides important information that can be used to understand the LULC dynamics and to monitor human activities in Brazilian biomes.

Furthermore, the data cube provides a database to support the scientific community with a wide range of applications, such as monitoring deforestation, calculating carbon emissions, monitoring and hydrological modelling, among others.

8. ACKNOWLEDGEMENTS

This work was funded by the Amazon Fund through the financial collaboration of the Brazilian Development Bank (BNDES) and the Foundation for Science, Technology and Space Applications (FUNCATE) no. 17.2.0536.1 (Brazil Data Cube project). This work is also supported by Group on Earth Observations (GEO) through the Earth Observation Cloud Credits Programme.

We also thank the Brazil Data Cube project team: Lúbia Vinhas, Ricardo Cartaxo, Thales Sehn Körting, Rennan Marujo, Jefferson de Souza Arcanjo, Raphael William da Costa, and Carlos Alberto Noronha.

REFERENCES

- Aghabozorgi, S., Seyed Shirshorshidi, A., Ying Wah, T., 2015. Time-Series Clustering - A Decade Review. *Inf. Syst.*, 53(C), 16–38.
- Almeida, C., Coutinho, A., Esquerdo, J., Adami, M., Venturieri, A., Diniz, C., Dessay, N., Durieux, L., Gomes, A., 2016. High spatial resolution land use and land cover mapping of the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data. *Acta Amazonica*, 46, 291 - 302.
- Appel, M., Pebesma, E., 2019. On-Demand Processing of Data Cubes from Satellite Image Collections with the gdalcubes Library. *Data*, 4(3), 92.
- Ariza-Porras, C., Bravo, G., Villamizar, M., Moreno, A., Castro, H., Galindo, G., Cabera, E., Valbuena, S., Lozano, P., 2017. CDCol: A Geoscience Data Cube that Meets Colombian Needs. A. Solano, H. Ordoñez (eds), *Advances in Computing*, 735, Springer International Publishing, Cham, 87–99.
- Asmaryan, S., Muradyan, V., Tepanosyan, G., Hovsepian, A., Saghatelyan, A., Astsatryan, H., Grigoryan, H., Abrahamyan, R., Guigoz, Y., Giuliani, G., 2019. Paving the Way towards an Armenian Data Cube. *Data*, 4(3), 117.
- Azevedo, T., S., Souza, C. M., J., Shimbo, J., Alencar, A., 2018. MapBiomass initiative: Mapping annual land cover and land use changes in Brazil from 1985 to 2017. *AGU Fall Meeting Abstracts*, 2018, B22A–04.
- Bravo, G., Castro, H., Moreno, A., Ariza-Porras, C., Galindo, G., Cabrera, E., Valbuena, S., Lozano-Rivera, P., 2017. Architecture for a Colombian Data Cube Using Satellite Imagery for Environmental Applications. A. Solano, H. Ordoñez (eds), *Advances in Computing*, 735, Springer International Publishing, Cham, 227–241.
- Brazilian Ministry of Environment, 2019. Brazilian biomes. <http://www.mma.gov.br/biomass/>. (Accessed on November, 2019).
- Brazil's National Institute for Space Research – INPE, 2020. Monitoring of the Brazilian Amazon Deforestation by Satellite - PRODES. <https://www.obt.inpe.br/prodes/>. (Accessed on 15 January 2020).
- Breiman, L., 2001. Random forests. *Machine learning*, 45(1), 5–32.
- Brooke, B., Lymburner, L., Lewis, A., 2017. Coastal Dynamics of Northern Australia – Insights from the Landsat Data Cube. *Remote Sensing Applications: Society and Environment*, 8, 94–98.
- Camara, G., Simoes, R., Andrade, P. R., Maus, V., Sánchez, A., Assis, L. F. F. G., Santos, L., Carvalho, A. Y., Maciel, A. M., Vinhas, L., de Queiroz, G. R., 2018. sits: version 1.12.5. <https://doi.org/10.5281/zenodo.1974065>.
- Chaves, M., Alves, M., Oliveira, M., Sáfyadi, T., 2018. A Geostatistical Approach for Modeling Soybean Crop Area and Yield Based on Census and Remote Sensing Data. *Remote Sensing*, 10(5), 680.
- Cloud Optimized GeoTIFF, 2019. Cloud Optimized GeoTIFF (COG): An imagery format for cloud-native geospatial processing. <https://www.cogeo.org/>. (Accessed on November, 2019).
- Committee on Earth Observation Sciences (CEOS), n.d. The Open Data Cube Initiative. https://docs.wixstatic.com/ugd/f9d4ea_1aea90c5bb7149c8a730890c0f791496.pdf. (Accessed on December, 2019).
- Dhu, T., Dunn, B., Lewis, B., Lymburner, L., Mueller, N., Telfer, E., Lewis, A., McIntyre, A., Minchin, S., Phillips, C., 2017. Digital Earth Australia – Unlocking New Value from Earth Observation Data. *Big Earth Data*, 1(1-2), 64–74.
- Dhu, T., Giuliani, G., Juárez, J., Kavvada, A., Killough, B., Merodio, P., Minchin, S., Ramage, S., 2019. National Open Data Cubes and Their Contribution to Country-Level Development Policies and Practices. *Data*, 4(4).
- Dwyer, J. L., Roy, D. P., Sauer, B., Jenkerson, C. B., Zhang, H. K., Lymburner, L., 2018. Analysis Ready Data: Enabling Analysis of the Landsat Archive. *Remote Sensing*, 10(9).
- Frantz, D., 2019. FORCE—Landsat + Sentinel-2 Analysis Ready Data and Beyond. *Remote Sensing*, 11(9), 1124.
- Gitelson, A. A., Merzlyak, M. N., Lichtenthaler, H. K., 1996. Detection of Red Edge Position and Chlorophyll Content by Reflectance Measurements Near 700 nm. *Journal of Plant Physiology*, 148(3), 501 - 508.
- Giuliani, G., Chatenoux, B., De Bono, A., Rodila, D., Richard, J.-P., Allenbach, K., Dao, H., Peduzzi, P., 2017. Building an Earth Observations Data Cube: Lessons Learned from the Swiss Data Cube (SDC) on Generating Analysis Ready Data (ARD). *Big Earth Data*, 1(1-2), 100–117.
- Hamunyela, E., Reiche, J., Verbesselt, J., Herold, M., 2017. Using Space-Time Features to Improve Detection of Forest Disturbances from Landsat Time Series. *Remote Sensing*, 9(6).
- Hamunyela, E., Verbesselt, J., de Bruin, S., Herold, M., 2016. Monitoring Deforestation at Sub-Annual Scales as Extreme Events in Landsat Data Cubes. *Remote Sensing*, 8(8).
- Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., Hobart, G. W., 2018. Disturbance-Informed Annual Land Cover Classification Maps of Canada's Forested Ecosystems for a 29-Year Landsat Time Series. *Canadian Journal of Remote Sensing*, 44(1), 67–87.

- Houborg, R., McCabe, M. F., 2018. A Cubesat Enabled Spatio-Temporal Enhancement Method (CESTEM) Utilizing Planet, Landsat and MODIS Data. *Remote Sensing of Environment*, 209, 211–226.
- Huete, A., Liu, H., Batchily, K., van Leeuwen, W., 1997. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment*, 59(3), 440 - 451.
- Killough, B., 2018. Overview of the Open Data Cube Initiative. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, Valencia, 8629–8632.
- Killough, B., 2019. The Impact of Analysis Ready Data in the Africa Regional Data Cube. *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, Yokohama, Japan, 5646–5649.
- Klink, C., Machado, R., 2005. Conservation of the Brazilian cerrado. *Conservation Biology*, 19(3), 707-713.
- Lewis, A., Oliver, S., Lymburner, L., Evans, B., Wyborn, L., Mueller, N., Raevksi, G., Hooke, J., Woodcock, R., Sixsmith, J., Wu, W., Tan, P., Li, F., Killough, B., Minchin, S., Roberts, D., Ayers, D., Bala, B., Dwyer, J., Dekker, A., Dhu, T., Hicks, A., Ip, A., Purss, M., Richards, C., Sagar, S., Trenham, C., Wang, P., Wang, L.-W., 2017. The Australian Geoscience Data Cube — Foundations and Lessons Learned. *Remote Sensing of Environment*, 202, 276–292.
- Lucas, R., Mueller, N., Siggins, A., Owers, C., Clewley, D., Bunting, P., Kooymans, C., Tissot, B., Lewis, B., Lymburner, L., Metternicht, G., 2019. Land Cover Mapping Using Digital Earth Australia. *Data*, 4(4), 143.
- 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.
- Metternicht, G., 2003. Vegetation indices derived from high-resolution airborne videography for precision crop management. *International Journal of Remote Sensing*, 24(14), 2855-2877.
- Nativi, S., Mazzetti, P., Craglia, M., 2017. A View-Based Model of Data-Cube to Support Big Earth Data Systems Interoperability. *Big Earth Data*, 1(1-2), 75–99.
- Nativi, S., Mazzetti, P., Santoro, M., Papeschi, F., Craglia, M., Ochiai, O., 2015. Big Data Challenges in Building the Global Earth Observation System of Systems. *Environmental Modelling & Software*, 68, 1–26.
- Picoli, M. C. A., Camara, G., Sanches, I., Simões, R., Carvalho, A., Maciel, A., Coutinho, A., Esquerdo, J., Antunes, J., Beggotti, R. A., Arvor, D., Almeida, C., 2018. Big earth observation time series analysis for monitoring Brazilian agriculture. *ISPRS Journal of Photogrammetry and Remote Sensing*, 1–12.
- Picoli, M. C. A., Rorato, A., Leitão, P., Camara, G., Maciel, A., Hostert, P., Sanches, I. D., 2020. Impacts of Public and Private Sector Policies on Soybean and Pasture Expansion in Mato Grosso—Brazil from 2001 to 2017. *Land*, 9(1).
- Pinty, B., Verstraete, M., 1992. GEMI: a non-linear index to monitor global vegetation from satellites. *Vegetatio*, 15–20.
- Rouse, R. W. H., Haas, J. A. W., Deering, D. W., 1974. Monitoring Vegetation Systems in the Great Plains with ERTS. *Third Earth Resources Technology Satellite-1 Symposium - Volume I: Technical Presentations*. NASA SP-351, NASA, 309–317.
- Santos, L., Ferreira, K. R., Picoli, M., Camara, G., 2020. Self-organizing maps in earth observation data cubes analysis. A. Vellido, K. Gibert, C. Angulo, J. D. Martín Guerrero (eds), *Advances in Self-Organizing Maps, Learning Vector Quantization, Clustering and Data Visualization*, Springer International Publishing, Cham, 70–79.
- Schwieder, M., Leitão, P. J., Bustamante, M. M. C., Ferreira, L. G., Rabe, A., Hostert, P., 2016. Mapping Brazilian savanna vegetation gradients with Landsat time series. *International Journal of Applied Earth Observation and Geoinformation*, 52, 361–370.
- Simoes, R., Picoli, M. C. A., Camara, G., Maciel, A., Santos, L., Andrade, P. R., Sánchez, A., Ferreira, K., Carvalho, A., 2020. Land use and cover maps for Mato Grosso State in Brazil from 2001 to 2017. *Scientific Data*, 7(1).
- Soterroni, A. C., Ramos, F. M., Mosnier, A., Fargione, J., Andrade, P. R., Baumgarten, L., Pirker, J., Obersteiner, M., Kraxner, F., Câmara, G., Carvalho, A. X. Y., Polasky, S., 2019. Expanding the Soy Moratorium to Brazil's Cerrado. *Science Advances*, 5(7).
- Spatio Temporal Asset Catalog, 2019. Spatio Temporal Asset Catalog (STAC) specification - making geospatial assets openly searchable and crawlable. <https://github.com/radiantearth/stac-spec>.
- Wiens, T. S., Dale, B. C., Boyce, M. S., Kershaw, G. P., 2008. Three way k-fold cross-validation of resource selection functions. *Ecological Modelling*, 212(3), 244–255.
- Xi, W., Du, S., Wang, Y. C., Zhang, X., 2019. A Spatiotemporal Cube Model for Analyzing Satellite Image Time Series: Application to Land-Cover Mapping and Change Detection. *Remote Sensing of Environment*.
- Yao, X., Liu, Y., Cao, Q., Li, J., Huang, R., Woodcock, R., Paget, M., Wang, J., Li, G., 2018. China Data Cube (CDC) for Big Earth Observation Data: Lessons Learned from the Design and Implementation. *2018 International Workshop on Big Geospatial Data and Data Science (BGDDS)*, IEEE, Wuhan, China, 1–3.
- Zhong, L., Hu, L., Yu, L., Gong, P., Biging, G. S., 2016. Automated mapping of soybean and corn using phenology. *ISPRS Journal of Photogrammetry and Remote Sensing*, 119, 151–164.