Brazil Data Cube Cloud Coverage (BDC^3) Viewer

Felipe Rafael de Sá Menezes Lucena, Elton Vicente Escobar-Silva, Rennan de Freitas Bezerra Marujo, Matheus Cavassan Zaglia, Lubia Vinhas, Karine Reis Ferreira and Gilberto Ribeiro de Queiroz

Earth Observation and Geoinformatics Division, National Institute for Space Research (INPE), São José dos Campos – SP – 12227-010 – Brazil

{felipe.lucena, elton.silva, rennan.marujo, matheus.zaglia, lubia.vinhas, karine.ferreira, gilberto.queiroz}@inpe.br

Abstract. *Remotely sensed Earth Observation (EO) data have exceeded a large scale and are increasingly available for different communities and thematic applications. In this matter, EO Data Cubes (EODC) are a promising solution to efficiently manage big EO data, enabling its access, processing and analysis. As final products, EODC provides analysis-ready data (ARD) for vegetation and land use and land cover (LULC) studies, for environmental monitoring urban growth studies, among others. One of the most used pre-selection criteria for image retrieval of big EO data is the average cloud cover statistic. This paper describes our initiative in the implementation of an on-demand view-based tool for cloud coverage embedded on the Brazil Data Cube (BDC) project based on the SpatioTemporal Asset Catalog (STAC).*

1. Introduction

In the last decade, remotely sensed Earth Observation (EO) data have exceeded the petabyte-scale milestone [Giuliani et al. 2019] and are increasingly available in diverse free and open access repositories [Giuliani et al. 2017]. This highlighted significant issues regarding storage, organization, management, and EO data analysis due to its volume, velocity, and variety [Giuliani et al. 2019]. However, to exploit the potential benefits of big EO data users are often required to have a high level of expertise and are frequently hampered by engaging in exhausting and tedious processes to discover data, and extract information and knowledge [Plag & Jules-Plag 2019].

Owing to the limitations of traditional acquisition, management, distribution, and analysis approaches of the big amount of satellite EO data available, exploitation of the full information potential of big EO data has not been achieved so far [Giuliani et al. 2019] and is at a lower level than desirable and feasible [Plag & Jules-Plag 2019]. Data size, heterogeneity, and complexity are the most notably limitations [Giuliani et al. 2019]. Therefore, new solutions are dearly needed to fulfill current analysis demands in this fast-changing field.

In this context, EO Data Cubes (EODC) are a promising solution to manage efficiently and effectively big EO data from different data repositories [Baumann et al. 2019; Nativi et al. 2017; Giuliani et al. 2017]. EODC are aimed to be a solution to store, organize, manage, and analyze large amounts of multi-sensor EO data [Poussin et al. 2019], as well as to provide access to EO Analysis-Ready Data (ARD) [Dwyer et al. 2018]. In other words, they are designed to harness big EO data, facilitating the access and use of ARD for immediate analysis in applications and for time-series exploitation

[Poussin et al. 2019]. As final products, EODC provide ARD for vegetation, land use and land cover (LULC), environmental monitoring and urban growth studies, among others [Ferreira et al. 2020; Giuliani et al. 2019; Poussin et al. 2019; Dhu et al. 2019].

One of the most used pre-selection criteria for image retrieval of big EO data is the average cloud cover statistic [Augustin et al. 2019]. This is because, for remote sensing of the earth's surface via optical sensors, some elements such as clouds act as no interest targets, which usually make the surface absent in some portion of the images [Zhong et al. 2017].

In this context, this paper describes our ongoing work to implement a novel viewer tool for cloud coverage embedded on the Brazil Data Cube (BDC) project, the Brazil Data Cube Cloud Coverage (BDC^3) viewer. Descriptions of the BDC project and BDC^3 viewer tool are presented in the following sections.

2. Brazil Data Cube (BDC) project

Since 2019, the Brazilian National Institute for Space Research (INPE) has been working in the BDC project (see http://brazildatacube.org/), which aims (i) to process remote sensing images of medium spatial resolution (10 to 30 meters) of the entire Brazilian territory into ARD datasets and assembling them as multidimensional cubes with at least three dimensions (space, time and spectral properties) [Ferreira et al. 2020; Picoli et al. 2020]; (ii) to propose and develop innovative methods and techniques to store, process and analyze EO data cubes using satellite image time series analysis (TSA), image processing and machine learning methods; (iii) to use the data cubes and methods to improve the generation of land use and cover change (LUCC) information for Brazil [Ferreira et al. 2020].

According to the Committee on Earth Observation Satellites (CEOS), ARD are "satellite data that have been processed to a minimum set of requirements and organized into a form that allows immediate analysis with a minimum of additional user effort and interoperability both through time and with other datasets" [Killough 2016]. In other words, ARD products can be described as data that have been processed to minimize the time and scientific knowledge required of users in such a way that allows analysis with a minimum of additional user effort [Dwyer et al. 2018].

The BDC project is creating ARD image collections and multidimensional data cubes from medium spatial resolution images of the sensors OLI/Landsat-8, MSI/Sentinel2 (A and B), and WFI/CBERS-4 [Picoli et al. 2020; Ferreira et al. 2020]. The latest is the fourth China-Brazil Earth Resources Satellite (CBERS-4) with the Wide-field Image camera (WFI), which produces medium resolution images in both visible and infrared wavelengths of the electromagnetic spectrum [Picoli et al. 2020]. Generally, in a data cube (DC) generation process, data acquisition and preprocessing are the first steps required. In BDC, all available images from the three repositories mentioned above are queried from their providers. To do so, a grid is used to search all available images of a specific grid cell on a given image collection. Then, these images are merged, reprojected, resampled, and gridded to a common spatial reference. Finally, a temporal compositing function is used to build regular intervals (e.g. 16 days or monthly, for instance) and reduce the data dimensionality according to a composition function (such as median or best quality pixel) (Figure 1) [Ferreira et al. 2020].

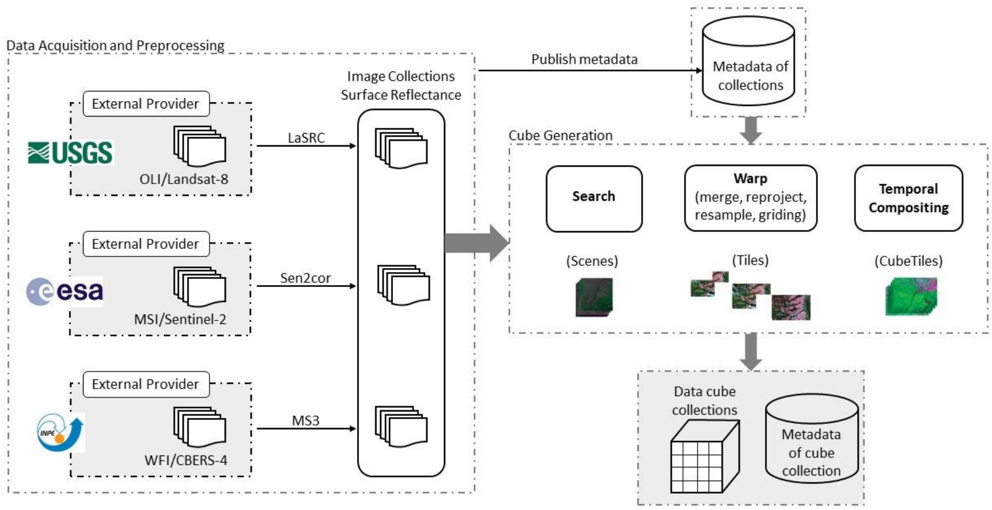

Figure 1: A summary of Brazil Data Cube (BDC) generation process. Adapted from Ferreira et al. (2020).

After the acquisition and preprocessing of OLI/Landsat-8, MSI/Sentinel (2A and 2B) and WFI/CBERS-4, their metadata are stored in an internal database catalog, then processed locally to generate the surface reflectance products through LaSRC [Vermote et al. 2016] and Sen2cor [Louis et al. 2016] atmospheric correction for OLI/Landsat-8 and MSI/Sentinel, respectively. To result in surface reflectance products, visibility masks are generated for each scene and data on the percentage of cloud coverage present in the image is cataloged in the image metadata.

3. Brazil Data Cube Cloud Coverage (BDC^3) Viewer

The BDC³ viewer is a tool to graphically visualize information about cloud cover in EODC, based on the SpatioTemporal Asset Catalog (STAC) specification (see https://github.com/radiantearth/stac-spec). The BDC³ viewer is being implemented in the BDC project using the BDC–STAC service. The BDC-STAC service implements the STAC specification, which defines how metadata of geospatial data are organized, consulted, and made available to users [Zaglia et al. 2019]. Therefore, the service aims cataloging and providing access to BDC metadata. A brief description of the BDC-STAC and BDC³ architecture is shown in Figure 2.

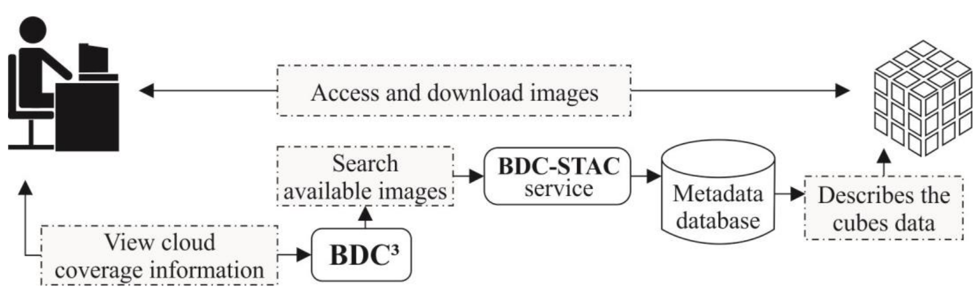

Figure 2: A description of BDC³ view-based tool and BDC-STAC service architecture. Adapted from Zaglia et al. (2019).

The BDC³ viewer proposes an extension of BDC-STAC's query possibilities for acquisition and visualization of cloud coverage information and will be able to answer questions such as *"Which data in my area of interest have less than 10% cloud cover?"* and "*In which year was the highest cloud coverage in December?"*. The tool allows user interaction, and all queries can be made per tile or polygon and are min/max coverage threshold-based. The BDC³ viewer provides four methods to get information about cloud cover from EO data cubes:

(i) seasonal (e.g. monthly) cloud coverage average;

(ii) total annual cloud coverage;

(iii) scene, area or period with maximum or minimum cloud coverage;

(iv) cloud coverage timeseries.

The seasonal average corresponds to the temporal grouping (average) of values for one or more tiles. For example, the monthly average represents the average between the cloud coverage values for each month. The total annual cloud coverage returns all scene coverage values for the queried period in a scatter plot. The selection of maximum and minimum queries the scene, area or period of interest according to the parameters defined by the user. Finally, the cloud coverage time-series query allows the display of coverage values for a given tile within the selected period.

The main idea of the tool is basically to show information about cloud cover in EODC as graphs and tables through an interactive web interface according to the defined parameters. Besides the visual information, the grouped data that generate graphics will be available for download to users. Two examples of possible queries to the metadata database through BDC³ viewer are shown as visual information in Figures 3 and 4. As a result, the users will be able to view information and acquire data on cloud coverage for a spatial extent of interest directly from the web without having to get images files or run complex algorithms.

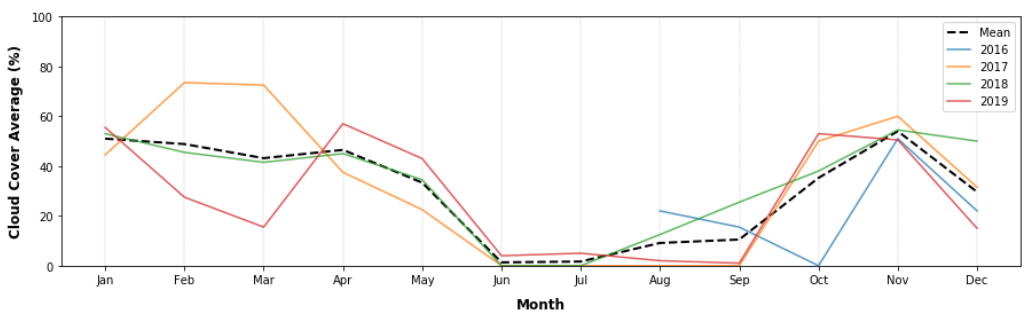

Figure 3: An example of a query-based on the monthly cloud coverage on BDC³. Just for illustration, it was chosen the cloud coverage values of the Landsat Path/Row 222068 from 2016 to 2019.

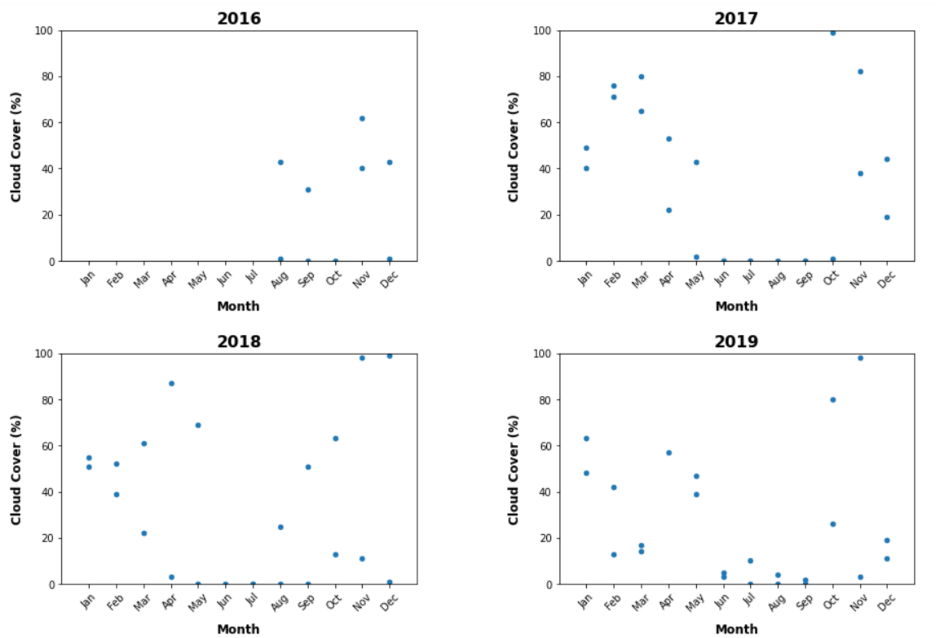

Figure 4: An example of a query-based on the cloud coverage absolute values by month on BDC³. Just for illustration, it was the chosen cloud coverage values of the Landsat Path/Row 222068 from 2016 to 2019.

4. Final considerations

As a work in progress, this paper aims to present the initial stages of the Brazil Data Cube Cloud Coverage (BDC^3) viewer, a novel view-based tool for cloud coverage embedded on the BDC project. BDC^3 viewer has a diversity of applications and its potential uses are the aid in scenes selection for agricultural monitoring, support in sensor selection for study in an area of interest, and the availability evaluation of scenes with clean pixels for a time interval, among others. As the interactive tool is already designed, the next steps are (i) complete the script and (ii) integrate the tool on the BDC as interactive maps.

5. References

AUGUSTIN, Hannah; SUDMANNS, Martin; TIEDE, Dirk; LANG, Stefan; BARALDI, Andrea. (2019). Semantic Earth observation data cubes. Data, 4(3), 102.

BAUMANN, Peter; MISEV, Dimitar; MERTICARIU, Vlad; HUU, Bang P. (2019). Datacubes: Towards space/time analysis-ready data. In Service-oriented mapping (pp. 269-299). Springer, Cham.

DHU, Trevor; GIULIANI, Gregory; JUÁREZ, Jimena; KAVVADA, Argyro; KILLOUGH, Brian; MERODIO, Paloma; MINCHIN, Stuart; RAMAGE, Steven. (2019). National open data cubes and their contribution to country-level development policies and practices. Data, 4(4), 144.

DWYER, John L.; ROY, David P.; SAUER, Brian; JENKERSON, Calli B.; ZHANG, Hankui K.; LYMBURNER, Leo. (2018). Analysis ready data: enabling analysis of the Landsat archive. Remote Sensing, 10(9), 1363.

FERREIRA, Karine. R.; QUEIROZ, Gilberto R.; CAMARA, Gilberto; SOUZA, Ricardo C. M.; VINHAS, Lúbia; MARUJO, Rennan F. B.; SIMOES, Rolf E. O.; NORONHA, Carlos A. F.; COSTA, Raphael W.; ARCANJO, Jeferson S.; GOMES, Vitor C. F; ZAGLIA, Matheus C. (2020). Using Remote Sensing Images and Cloud Services on Aws to Improve Land Use and Cover Monitoring. In 2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS) (pp. 558-562). IEEE.

GIULIANI, Gregory; CHATENOUX, Bruno; BONO, Andrea D.; RODILA, Denisa; RICHARD, Jean-Philippe; ALLENBACH, Karin; DAO, Hy; PEDUZZI, Pascal. (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), pp. 100–117.

GIULIANI, Gregory; CAMARA, Gilberto; KILLOUGH, Brian; MINCHIN, Stuart. (2019). Earth observation open science: Enhancing reproducible science using data cubes.

KILLOUGH, Brian. (2016). CEOS land surface imaging analysis ready data (ARD) description document.

NATIVI, Stefano; MAZZETTI, Paolo; CRAGLIA, Max. (2017). A view-based model of data-cube to support big earth data systems interoperability. Big Earth Data, 1(1-2), 75-99.

PICOLI, Michelle C. A.; SIMOES, Rolf; CHAVES, Michel; SANTOS, Lorena A.; SANCHEZ, Alber; SOARES, Anderson; SANCHES, Ieda D.; FERREIRA, Karine R.; QUEIROZ, Gilberto R. (2020). CBERS DATA CUBE: A POWERFUL TECHNOLOGY FOR MAPPING AND MONITORING BRAZILIAN BIOMES. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 5(3).

PLAG, Hans-Peter; JULES-PLAG, Shelley A. (2019). A Transformative Concept: From Data Being Passive Objects to Data Being Active Subjects. Data, 4(4), 135.

POUSSIN, Chartlotte; GUIGOZ, Yaniss; PALAZZI, Elisa; TERZAGO, Silvia; CHATENOUX, Bruni; GIULIANI, Gregory. (2019). Snow Cover Evolution in the Gran Paradiso National Park, Italian Alps, Using the Earth Observation Data Cube. Data, 4(4), 138.

ZAGLIA, Matheus C.; VINHAS, Lubia; QUEIROZ, Gilberto R. de; SIMOES, Rolf. (2019). Catalogação de Metadados do Cubo de Dados do Brasil com o SpatioTemporal Asset Catalog. GEOINFO, 20 Years After!, 280.

ZHONG, Bo; CHEN, Wuhan; WU, Shanlong; HU, Longfei; LUO, Xiaobo; IU, Qinhuo. (2017). A Cloud Detection Method Based on Relationship Between Objects of Cloud and Cloud-Shadow for Chinese Moderate to High Resolution Satellite Imagery. In IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (p. 4898-4908). IEEE.