















Article

Earth Observation Data Cubes for Brazil: Requirements, Methodology and Products

Karine R. Ferreira ^{*}, Gilberto R. Queiroz , Lúbia Vinhas , Rennan F. B. Marujo , Rolf E. O. Simoes , Michelle C. A. Picoli , Gilberto Camara , Ricardo Cartaxo, Vitor C. F. Gomes, Lorena A. Santos , Alber H. Sanchez , Jeferson S. Arcanjo, José Guilherme Fronza , Carlos Alberto Noronha, Raphael W. Costa, Matheus C. Zaglia, Fabiana Zioti , Thales S. Korting , Anderson R. Soares , Michel E. D. Chaves  and Leila M. G. Fonseca

Earth Observation and Geoinformatics Division, National Institute for Space Research (INPE), Avenida dos Astronautas, 1758, Jardim da Granja, São José dos Campos, SP 12227-010, Brazil; gilberto.queiroz@inpe.br (G.R.Q.); lubia.vinhas@inpe.br (L.V.); rennan.marujo@inpe.br (R.F.B.M.); rolf.simoes@inpe.br (R.E.O.S.); michelle.picoli@inpe.br (M.C.A.P.); gilberto.camara@inpe.br (G.C.); ricardo.cartaxo@inpe.br (R.C.); vitor.gomes@inpe.br (V.C.F.G.); lorena.santos@inpe.br (L.A.S.); alber.ipia@inpe.br (A.H.S.); jeferson.arcanjo@inpe.br (J.S.A.); jose.fronza@inpe.br (J.G.F.); carlos.noronha@inpe.br (C.A.N.); raphael.costa@inpe.br (R.W.C.); matheus.zaglia@inpe.br (M.C.Z.); fabiana.zioti@inpe.br (F.Z.); thales.korting@inpe.br (T.S.K.); anderson.soares@inpe.br (A.R.S.); michel.chaves@inpe.br (M.E.D.C.); leila.fonseca@inpe.br (L.M.G.F.)

* Correspondence: karine.ferreira@inpe.br

Received: 4 November 2020; Accepted: 4 December 2020; Published: 9 December 2020



Abstract: Recently, remote sensing image time series analysis has been widely used to investigate the dynamics of environments over time. Many studies have combined image time series analysis with machine learning methods to improve land use and cover change mapping. In order to support image time series analysis, analysis-ready data (ARD) image collections have been modeled and organized as multidimensional data cubes. Data cubes can be defined as sets of time series associated with spatially aligned pixels. Based on lessons learned in the research project e-Sensing, related to national demands for land use and cover monitoring and related to state-of-the-art studies on relevant topics, we define the requirements to build Earth observation data cubes for Brazil. This paper presents the methodology to generate ARD and multidimensional data cubes from remote sensing images for Brazil. We describe the computational infrastructure that we are developing in the Brazil Data Cube project, composed of software applications and Web services to create, integrate, discover, access, and process the data sets. We also present how we are producing land use and cover maps from data cubes using image time series analysis and machine learning techniques.

Keywords: analysis-ready data; data cubes; image time series analysis; machine learning; land use and cover mapping

1. Introduction

The growing demands for natural resources and the need to deal with climate change's effects resulting from greenhouse gas emissions present significant challenges for our societies. To understand how these changes are affecting the environment, one of our primary sources of information is land use and cover change data obtained by Earth observation satellites [1]. Analysis of remote sensing images is the most effective way to produce up-to-date information about the Earth's surface status [2].

Nowadays, researchers have free access to an unprecedentedly large number of remote sensing images collected by different satellites and sensors with distinct spatial, temporal, and spectral resolutions.

In 2019, five petabytes of images were produced by the satellites Landsat-7/8, CBERS-4/4A, Terra/Aqua, and Sentinel-1/2/3 [3]. Big Earth observation data sets bring new challenges and opportunities, including the novel generation of technological solutions to store, process, disseminate, and analyze them [4]. One of these opportunities is the use of time series analysis to extract landscape change information from many remote sensing images [2,5].

Remote sensing satellites revisit the same place over the Earth's surface at different times and produce image sequences of the same location over time. Time series derived from these sequences are very useful to investigate the dynamics of the environment over time. Image time series analysis and machine learning methods have been widely used for land use and cover classification and change detection with good results [6–11].

To better support time series analysis, researchers produce analysis-ready data (ARD) from satellite imagery and organize the data as multidimensional data cubes [12]. Siqueira et al. [13] defines ARD as “satellite data 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 data sets.” ARD result from processing satellite imagery from data acquisition to radiometric calibration, and through additional conversions, to top-of-atmosphere (TOA) reflectance, and finally surface reflectance [14]. The term data cube refers to a set of image time series associated with spatially aligned pixels [15]. Each element of an Earth observation (EO) data cube has two spatial dimensions and one temporal dimension, and is associated with a set of values.

Recent initiatives to create EO data cubes from remote sensing images for specific regions include the Australian Data Cube [12], the Swiss Data Cube [14], the Armenian Data Cube [16], the Catalan Data Cube [17], and the Africa Regional Data Cube [18]. Meanwhile, novel software platforms have been developed to access, prepare, index, and handle large amounts of remote sensing images and data cubes, such as the Open Data Cube (ODC) [12], e-Sensing [19], and the Google Earth Engine (GEE) [20]. These initiatives and platforms are paving the way for broadening the use of EO data to larger communities of users, supporting decision-makers with information converted into meaningful geophysical variables, and ultimately unlocking the information power of EO data [21].

This paper presents the main requirements to build EO data cubes for Brazil and the methodology to generate ARD and data cubes from remote sensing images. We also describe the computational infrastructure that we are developing in the Brazil Data Cube project. To show the potential of the data sets and software products that we are producing, this paper presents the land use and cover maps created from the EO data cubes using image time series analysis and machine learning methods.

2. Requirements for EO Data Cubes for Brazil

The Brazilian National Institute for Space Research (INPE) is the leading government research organization working in Earth observation and space science areas. INPE is the main institute in Brazil responsible for the China–Brazil Earth Resources Satellite (CBERS) program, a cooperative project between Brazil and China [22]. Under this program, the most recent Earth observation satellite CBERS 04A was launched in December 2019.

Since 1988, INPE has led projects to produce official information about deforestation and land use and cover changes in Brazil. The DETER project provides deforestation alerts [23], and the PRODES project monitors clear-cut deforestation, providing accurate official deforestation rates in the Brazilian Amazon forest and other biomes [24]. The data sets produced by DETER and PRODES can be freely accessed through the TerraBrasilis platform [25]. The TerraClass project identifies and maps what the deforested areas by PRODES have become. The resulting maps can be used to evaluate the spatiotemporal and semantic land dynamics of these areas [26]. These projects are part of a forest monitoring system recognized by its impact on deforestation assessment and control initiatives [27].

Recently, INPE has worked in projects aiming to exploit the unprecedented remote sensing imagery freely available nowadays by using big data technologies, cloud computing environments, and machine learning methods. The e-Sensing project ran from 2015 to 2018, investigating new methods

and techniques to substantially improve the extraction of land use and cover change information from big Earth Observation data sets in an open and reproducible way. Based on the know-how acquired in the e-Sensing project and on national demands on land use and cover monitoring, the following requirements were defined to build EO data cubes for Brazil.

2.1. Image Time Series Analysis and Machine Learning

In the e-Sensing project, different classification methods, based on satellite image time series analysis and machine learning algorithms, were proposed, tested, and implemented. One of these methods is the time-weighted dynamic time warping (TWDTW) algorithm [28]. The authors improved the dynamic time warping (DTW) method for land use and cover classification, by introducing a time constraint to account for seasonality of land use types. This allows users to include information about the agricultural calendar in the region of interest, thereby obtaining better results.

Besides TWDTW, other machine learning methods, such as the support vector machine (SVM) and random forest (RF), were applied to classify satellite image time series for the Mato Grosso state and other regions in the Cerrado biome, Brazil [8,10,11]. These results presented good accuracy and made explicit the potential of image time series analysis and machine learning methods for land use and cover mapping. Thus, the main motivation to create multidimensional data cubes from ARD remote sensing images for all Brazilian territory is to be able to apply these methods for national land use and cover mapping. An important requirement is to continue investigating and advancing in techniques based on machine learning and image time series analysis, exploiting methods based not only on the temporal dimension but also on spatial context, using spatiotemporal heterogeneity and correlation [29].

2.2. ARD and Data Cubes of Medium-Resolution Satellite Images

In the e-Sensing project, we mainly used the MODIS sensor products MOD13Q1 and MYD13Q1 that have 250 m of spatial resolution and 16 days of temporal resolution. These products were created using a temporal composition method that selected the best pixels during each period of 16 days of the Terra and Aqua satellites. That increased the chances of finding values without cloud cover and enabled us to obtain dense time series.

The spatial resolution of the MOD13Q1 and MYD13Q1 products is not adequate for mapping agricultural activity in regions of smallholdings and identification of small deforestation areas. In our experiments, areas smaller than ~11 ha could not be classified accurately using MODIS data.

To improve the ability to identify small areas of land use and cover, an important requirement is to produce advancements in the use of medium-resolution images from CBERS-4 (64 m), Landsat-8 (30 m), and Sentinel-2 (10, 20, and 60 m) satellites. We need to create ARD image collections from the satellite imagery for all Brazilian territory and to model these collections as multidimensional data cubes to support time series analysis.

2.3. Land Use and Cover Samples and Data Sets

Based on our previous experiences, we confirmed that a classified map's accuracy depends directly on the quality of the training samples used by the machine learning methods. Thus, it is essential to advance the sample collection tools and methodology, and techniques used to assess and improve the quality of these samples, including feature selection and outlier removal [30].

Land use and cover samples are collected by different projects and individuals, using different methods, such as in situ gathering in fieldwork and visual interpretation of high-resolution satellite images. An important requirement is to be able to describe samples with proper metadata that characterize their differences and organize them in a shared database to facilitate the reproducibility of experiments. It is also important to develop tools to easily discover, query, access, and process this shared sample database.

2.4. Interoperability and Web Services

In the e-Sensing project, several technologies for data management, access, and processing were tested and used [19]. Initially, we organized and stored time series of data in binary files in the local file system, and developed proprietary technologies to access them. This solution proved inefficient when processing large volumes of data, especially when coordinating multiple simultaneous data access. To solve this problem, we adopted the SciDB open array database [31].

We also tested alternative solutions to SciDB, such as Apache Hadoop, and concluded that SciDB had better computational performance and a lower cost of adapting legacy software [19]. The SciDB solution had solved problems related to the coordination of data access. However, it presented a technical difficulty of interoperability with other systems, especially those developed under the e-Sensing project, such as the Satellite Image Time Series (sits) R package [32]. Thus, a fundamental requirement is to use open and consolidated technologies that facilitate interoperability with other systems.

One way to improve the interoperability between systems is to develop software components as Web services. Hence, we developed the Web Time Series Service (WTSS), a lightweight Web service to retrieve time series from big data sets in an open format and based on consolidated technologies [33]. WTSS proved to be suitable for dealing with big amounts of remote sensing images by providing server-side processing. The development of software components as Web services facilitates the integration with different programming languages, such as R, Python, and C++, and different types of client applications. It also enables the adoption of the moving code paradigm and server-side processing [34].

2.5. Cloud Computing and Distributed Processing Environments

Cloud computing and distributed processing environments have been supplied as services by different providers, such as Amazon Web Services (AWS) and Microsoft Azure Cloud Services. These environments provide development tools, storage, virtual machines, databases, and network services in a highly scalable way. Some providers also contain large repositories of Earth observation satellite image collections as open data, meaning that they can be used without restrictions and with minimum cost. As an example, the Open Data on AWS [35] program offers data buckets with Landsat-8, Sentinel-2, and CBERS-4 imagery.

Brazil is a huge country, with a territory of over 8,500,000 km². To build ARD and data cubes of Sentinel-2 (A and B) images from 2015 to 2020 for the entire country, we need to process 650 terabytes of data. Therefore, it is necessary to exploit the potential of these cloud computing environments to handle large amounts of remote sensing images efficiently. These environments do not impose any API for data processing, allowing the development of open source technologies in non-proprietary scripting languages, and thereby reducing potential vendor lock-in effects. We tested the cloud services on AWS to access remote sensing images stored as open data and concluded that they are a good alternative to generate and process EO data cubes [36].

3. Methodology for EO Data Cube Generation

Driven by the requirements described in Section 2, the Brazil Data Cube (BDC) project has four main objectives: (1) create ARD image collections from medium-resolution remote sensing images (10 to 64 m) for all Brazilian territory; (2) model these ARD images as multidimensional data cubes with three or more dimensions that include space, time, and spectral-derived properties, mainly to support image time series analysis; (3) use, propose, and develop big data technologies, such as cloud computing and distributed processing environments, to create, store, and process these data cubes; and (4) create land use and cover information, for Brazil, from these data cubes using satellite image time series analysis, machine learning methods, and image processing procedures. Figure 1 presents a general overview of the BDC project.

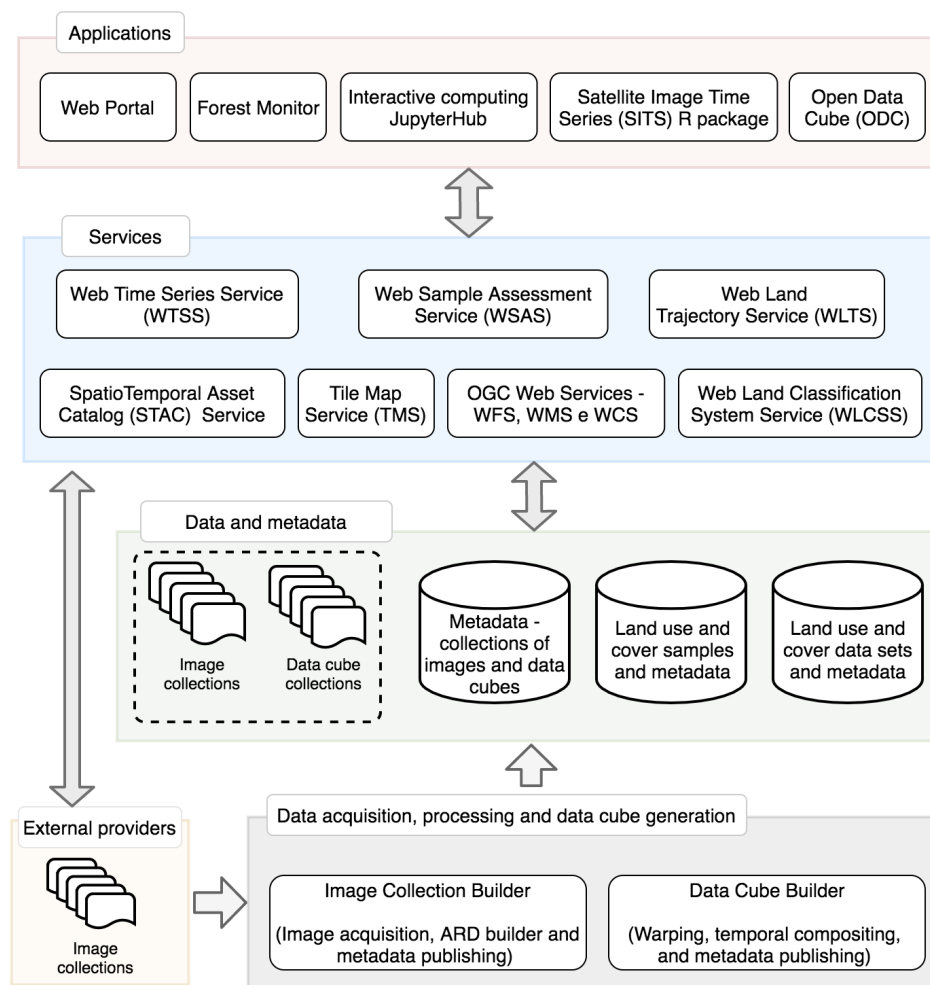


Figure 1. Brazil Data Cube Project: data and software products.

3.1. Data Acquisition and ARD Processing

Figure 2 details the process of data acquisition, the ARD procedure, and the data cube generation. Currently, we are using images from the Sentinel-2 (A and B) MultiSpectral Instrument (MSI), Landsat-8 Operational Land Imager (OLI), Thermal Infrared Sensor (TIRS), CBERS-4 Advanced Wide-Field Imager (AWFI), CBERS-4 Multispectral Camera (MUX), and AQUA/TERRA Moderate Resolution Imaging Spectroradiometer (MODIS), from 2016 to 2020, and covering all of Brazilian territory.

The images are collected from their different original providers and pre-processed into surface reflectance products. The atmospheric correction procedure depends on each sensor's characteristics, and different algorithms and software are used. For instance, for Sentinel-2/MSI and Landsat-8/OLI we use the LaSRC software [37] and for CBERS-4/AWFI and MUX the MS3 software [38]. For each image, cloud and cloud shadow masks are computed. For Sentinel-2 and Landsat-8, the Fmask (version 4.2) is being used [39] because it has presented the best overall accuracy, compared with MAJA, Sen2Cor, and s2cloudless, on images of the Amazon region in Brazil [40]. For CBERS-4 we use the CMASK algorithm because it does not have spectral bands need by Fmask. Once this process is performed and the resulting data set is cataloged, the surface reflectance image collections are used to generate the data cubes.

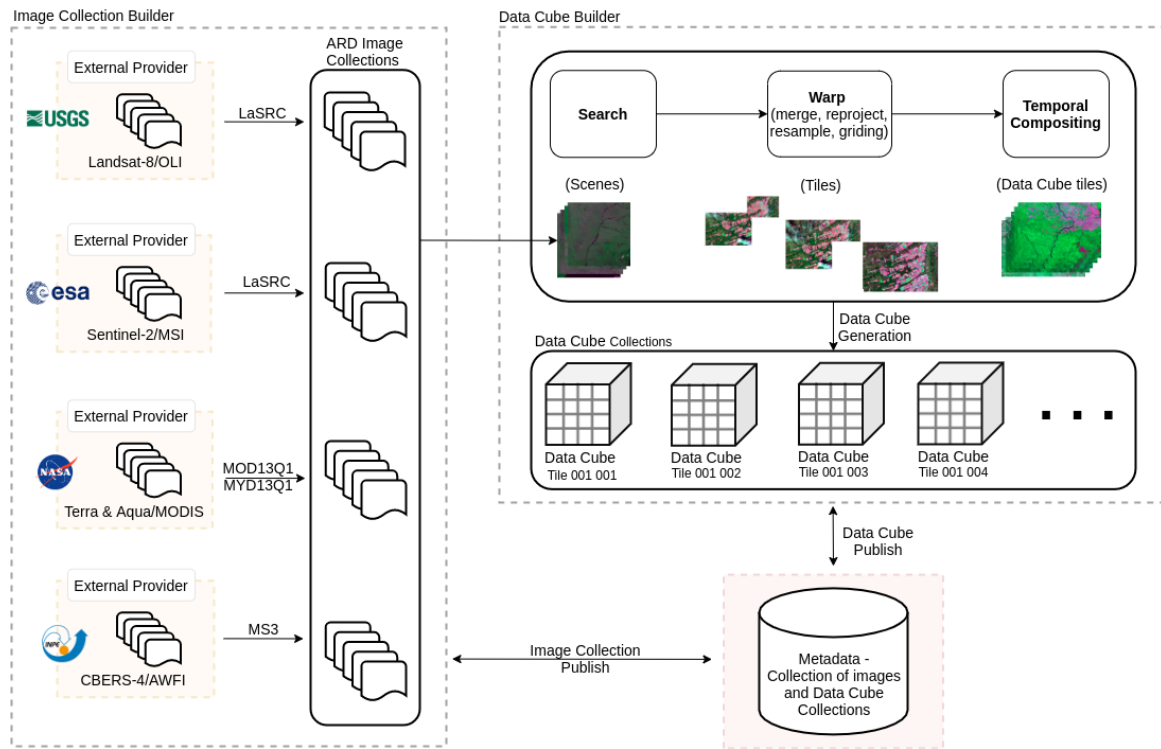


Figure 2. Data acquisition, ARD processing, and data cube generation.

3.2. Tiling System

The organization of data cubes is based on a spatial partition of the territory. This partition is represented by a grid of tiles where every pixel can be efficiently located within a tile. In the BDC project, the data cubes are created by tiles, and each tile of a grid generates a set of files in the disk. Each file is associated with a specific data cube and with a grid tile, and contains information about a specific property or band. Considering the spatial resolution of the images and aiming at maintaining files that can be easily manageable, three hierarchical grids were defined using an Albers equal area projection and SIRGAS 2000 datum. The three grids were generated by taking -54 longitude as the central reference and defining tiles of 6×4 , 3×2 , and 1.5×1 degrees, referred to as *BDC_LG* (large), *BDC_MD* (medium) and *BDC_SM* (small), as shown in Figure 3 and described in Table 1.

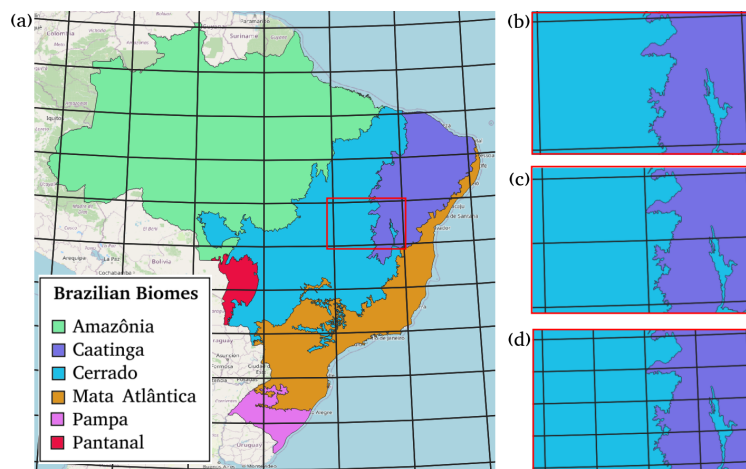


Figure 3. Hierarchical BDC tiling system showing *BDC_LG* overlaid on Brazilian Biomes (a), illustrating that one *BDC_LG* tile (b) contains four *BDC_MD* tiles (c) and that *BDC_MD* contains four *BDC_SM* tiles (d).

Table 1. Parameters of the hierarchical tiling system used in BDC.

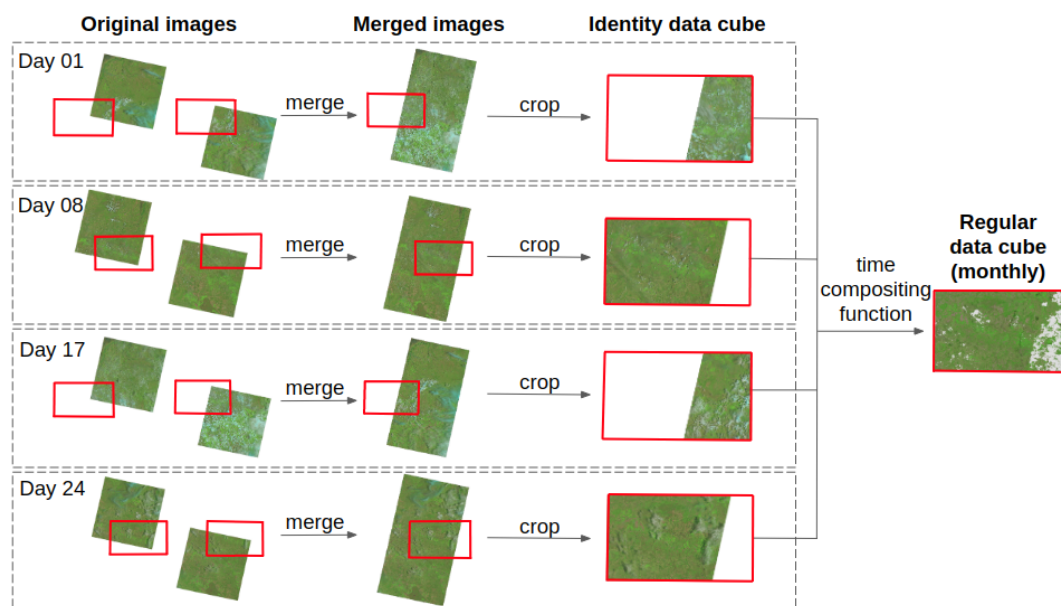
Grid	Collection	Extension	Image Size
BDC_LG	CBERS-4 AWF1 (64 m)	672 × 440 km	10,504 × 6865 pixels
BDC_MD	Landsat-8 OLI (30 m)	336 × 220 km	11,204 × 7324 pixels
BDC_SM	Sentinel-2 MSI (10 m)	168 × 110 km	16,806 × 10,986 pixels

3.3. Data Cube Generation

We are creating identity data cubes and regular data cubes. Currently, each data cube contains images from a single sensor. An identity data cube uses all available images in a period of time, tiled to a common grid, without applying a temporal compositing function, as illustrated in Figure 4. An identity data cube is not regular in the time dimension; it instead maintains all available images with their original acquisition dates.

The regular data cubes are created by using functions for temporal compositing. They are regular in the time dimension. These temporal compositing functions reduce the original time dimension to a longer unit (e.g., a month or 16 days), generating observations that are regularly spaced in time. Figure 4 shows the entire process to create a monthly regular data cube.

The process to generate data cubes illustrated in Figure 4 is composed of two steps: warping and temporal compositing. The warping process generates the identity data cubes. The input images are resampled to the same spatial resolution using a bilinear function for all bands and nearest neighbor for cloud masks. Then, the images are re-projected to the common projection, and images from the same day are merged into one mosaic. Finally, the mosaic is cropped into the grid tile extension.

**Figure 4.** The process to generate a monthly regular data cube, considering a grid tile.

Regular data cubes require the definition of a regular time step (e.g., one month or 16 days) that will guide the temporal compositing of multiples images available in each time step. After the warping process, the time compositing function is applied to the identity data cubes. It defines the approach to selecting or generating a value representing the series of observations available during the time step. It should be noticed that temporal compositing considers only valid observations, which are not detected as clouds, cloud shadows, or snows, nor does it account for the quality assessment of each band.

Currently, BDC supports three functions for temporal compositing: (i) median, (ii) average, and (iii) stack, as illustrated in Figure 5. The median compositing function consists of, given a time

step, for each image band, calculating the median value of the valid observations and using it as the observation for that time step. The median function uses the central value, while keeping an existing observation, removing outliers in a given time step. The average composition uses the mean of all valid observations in the time step. This also removes outliers, but in this case a new smoothed observation is generated.

The stack time compositing function is the best pixel approach; it consists of ranking the time step images by the number of valid observations and selecting the observation from the best-ranked image. The stack compositing ensures that observations coming from less cloud coverage have higher probabilities of being selected to represent the time step.

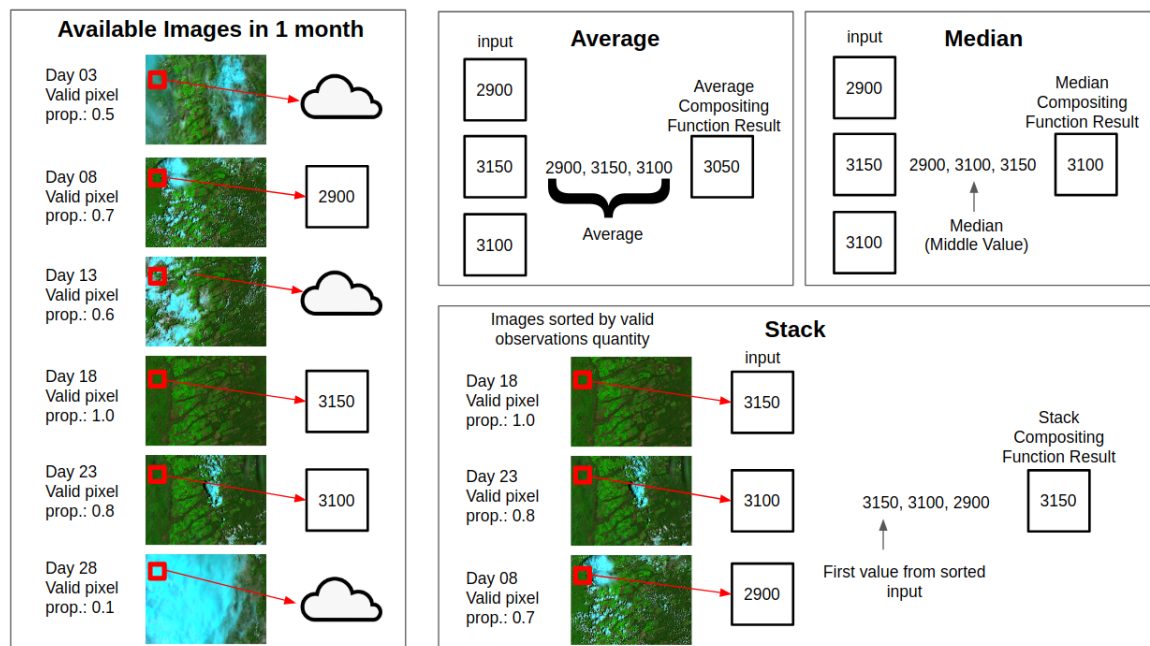


Figure 5. Brazil Data Cube time compositing functions (average, median, and stack) applied on a pixel obtained from a monthly time step.

The image collections and data cubes are distributed as Cloud Optimized GeoTIFF (COG) files, with an internal organization that enables efficient data access in the distributed and high-performance cloud environments. In the data cubes, each GeoTIFF file contains information of a band or property of a specific tile of the BDC tiling system and of a certain time period. Besides the spectral bands from original satellite images, the data cube files also contain: (1) cloud mask; (2) spectral indices (NDVI and EVI); (3) the total number of observations for each pixel before time compositing; (4) the number of valid observations (excluding cloud, cloud shadow, snow, or partial images) for each pixel before time compositing; (5) a data provenance band that indicates the image of origin of each pixel, when the stack compositing function is applied; (6) a thumbnail for each time step.

3.4. Data Cube Validation and Metadata

We are conducting data quality assessment routines over image collections (surface reflectance products) and data cube collections. For surface reflectance products, we evaluate radiometric and geometric characteristics, cloud mask algorithms [40], and uncertainties in pre-processing steps (e.g., comparing atmospheric correction algorithms, Sen2Cor to Sentinel-2A/2B surface reflectance products, and LaSRC to Landsat-8 OLI surface reflectance products). We are exploring data quality dimensions for the data cubes, such as data completeness, consistency, uniqueness, validity, and accuracy [41]. Additionally, we are looking into quality analysis (e.g., spatial redundancy

pixel counting), validation of the time compositing, checking the consistency of mask quality, date provenance of the observations, and differences caused by different versions of processing algorithms.

The metadata about the image collections and data cubes are stored in a relational database. We are using the SpatioTemporal Asset Catalog (STAC) [42] as the language to describe the metadata. STAC is an open specification, based on JSON and RESTful, that has its origins in 14 different organizations working together to increase the interoperability of searching for geospatial data, including satellite imagery. STAC also provides API mechanics to search for and access geospatial data, which are being used by the applications using the BDC data sets.

4. Software and Data Products

The software products that are being developed in the BDC project are free and open-source and can be grouped into two categories, services and applications, as shown in Figure 1. All project data sets and metadata are accessed and processed through Web services. Applications refer to software products for final users, including systems with graphical interfaces and APIs (application programming interfaces) for high-level programming languages such as R and Python. Web services provide the interfaces between the project data sets and applications.

4.1. Web Services

In the BDC project, we are developing Web service specifications defined by the project team, such as Web Time Series Service (WTSS) and Web Land Trajectory Service (WLTS), and by third-party entities, such as the Tile Map Service (TMS) by THE Open Source Geospatial Foundation; and the Web Map Service (WMS), Web Feature Service (WFS), and Web Coverage Service (WCS) by the Open Geospatial Consortium (OGC) and ISO [43]. Besides using standard Web services protocols defined by OGC and ISO to promote interoperability, we decided to develop other lightweight Web services tailored to specific purposes.

As described in Section 3.4, the image collections and data cubes metadata follow the STAC specification. Thus, the STAC service provides a standard interface that allows users and applications to query and discover all image collections and data cubes stored in the project databases. Besides the STAC service, we developed two clients based on the STAC API; the `stac.py` package in Python and the `rstac` package in R.

The WTSS is a lightweight Web service for handling time series data from remote sensing imagery. In this service, we use the term coverage to refer to image collections and multidimensional data cubes. Given a coverage, a spatial location, and a time interval, it retrieves the time series associated with them [33].

As shown in Figure 1, besides the image collections and data cubes, we are organizing land use and cover samples, and data sets and their metadata in database systems, and developing Web services to deal with them. These samples are heterogeneous. They are gathered by different specialists using specific land use and cover classification systems and distinct means such as in situ fieldwork and visual interpretation of high-resolution satellite images. All these details are stored as metadata in the database system.

We are also organizing data sets from distinct projects and institutions containing land use and cover information about the Brazilian territory, such as PRODES, DETER, and TerraClass. These data sets are also created by distinct groups using particular land use and cover classification systems. To deal with distinct classification systems associated with samples and data sets, we developed the Web Land Classification System Service (WLCSS). The WLCSS is responsible for describing and mapping distinct land use and cover classification systems.

Given a spatial location, the Web Land Trajectory Service (WLTS) returns how that location was classified by distinct land use and cover data sets over time. The term land trajectory refers to a sequence of land use and cover classes, ordered in time, associated with a spatial location. Figure 6 illustrates how the WLTS works. Based on the land use and cover data sets generated by TerraClass, PRODES, and DETER, this picture shows a land trajectory associated with a specific location in Para

(PA) state, Brazil. This location was classified as forest in 2012 by PRODES; then as deforestation alert in 2013 by DETER; then as deforestation in 2013 by PRODES; and as pasture in 2014 by TerraClass. This service is useful for helping users with the land use and cover sample gathering and with validation of the classification map.

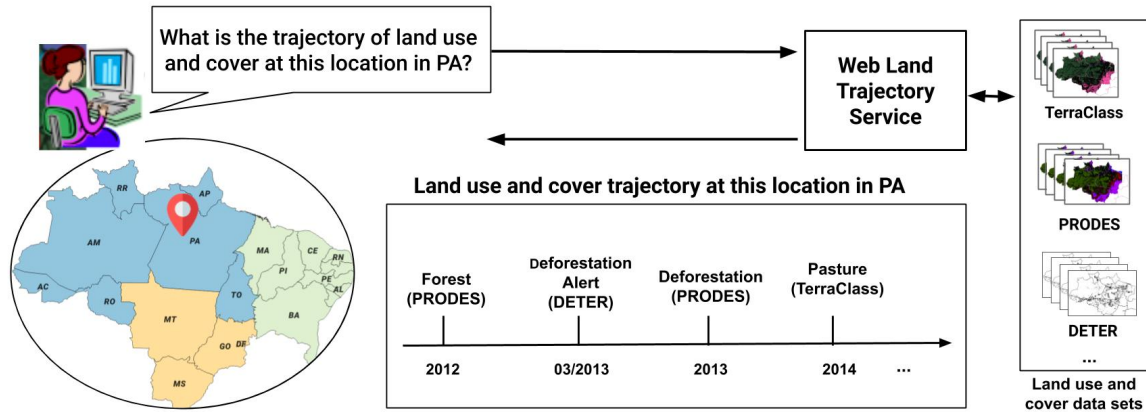


Figure 6. Web Land Trajectory Service (WLTS).

The Web Sample Assessment Service (WSAS) is being implemented to evaluate and improve the quality of land use and cover samples. These samples are used to train machine learning methods. They can be noisy, causing negative effects on classification performance. The accuracy of classified maps depends directly on the quality of the training samples used.

The WSAS is based on a method developed in the *sits R* package that identifies data wrongly labeled or data with low discrimination when mixed with other land use and cover classes. This method gets the time series associated with samples from data cubes and uses the self-organizing map (SOM) neural network and Bayesian inference to calculate the probability of each sample belonging to each class.

4.2. Applications

Through the BDC Web portal, shown in Figure 7, users can access, visualize using a temporal slider, download, and analyze differences between two time periods of image collection and those between two data cubes. The green grid tiles represent tiles that have ready data cubes associated with them. Figure 7 shows the Landsat-8 data cubes that are ready for Brazil, created by using the stack compositing function and a temporal step of 16 days.

The Open Data Cube (ODC) is a framework that provides tools for indexing, visualization, processing, and analysis of EO data. It is an open-source framework that is used as an infrastructure for the management of data cubes for a growing number of projects, such as the Australian Geoscience Data Cube [12], Swiss Data Cube [14], and Digital Earth Africa [18].

At INPE, we have an ODC instance with all image collections and data cubes created in the BDC project. To integrate ODC framework with BDC data products, we developed a tool called *stac2odc* for accessing our data sets and automatically indexing them in the ODC framework. This tool collects the necessary metadata from the BDC STAC Web service, converts them to the appropriate formats, and then records the data sets in the INPE's ODC instance. This ODC instance has direct access to the data sets produced in the BDC project.

In the BDC project, we are using four components of the ODC framework: (1) ODC Explorer to visualize the footprint and metadata of the BDC data products; (2) ODC Open Web Services to create the OGC WMS and WCS services for the BDC data sets; (3) ODC Python API in a JupyterHub environment to access and process the BDC data sets in Python; and (4) ODC Temporal Statistics Tools. Figure 8 shows the ODC Explorer and the ODC Python API with the CBERS4 data cubes created in the BDC project.

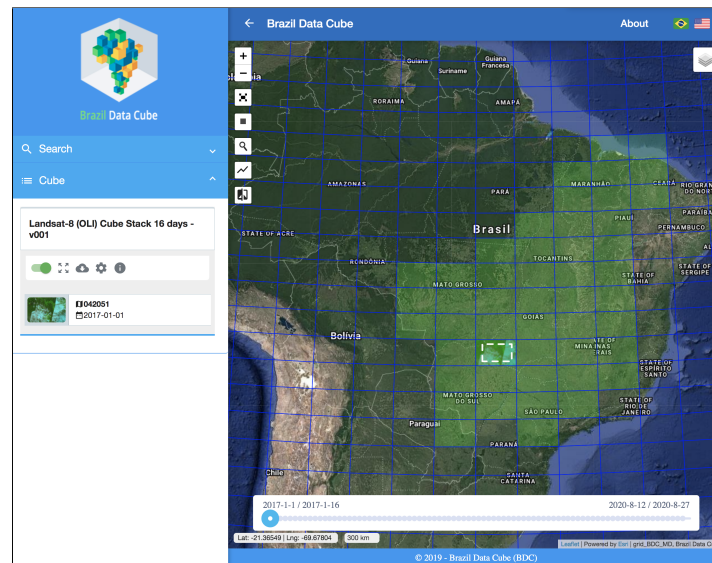


Figure 7. BDC Web portal available at: <http://brazildatacube.dpi.inpe.br/portal/explore>.

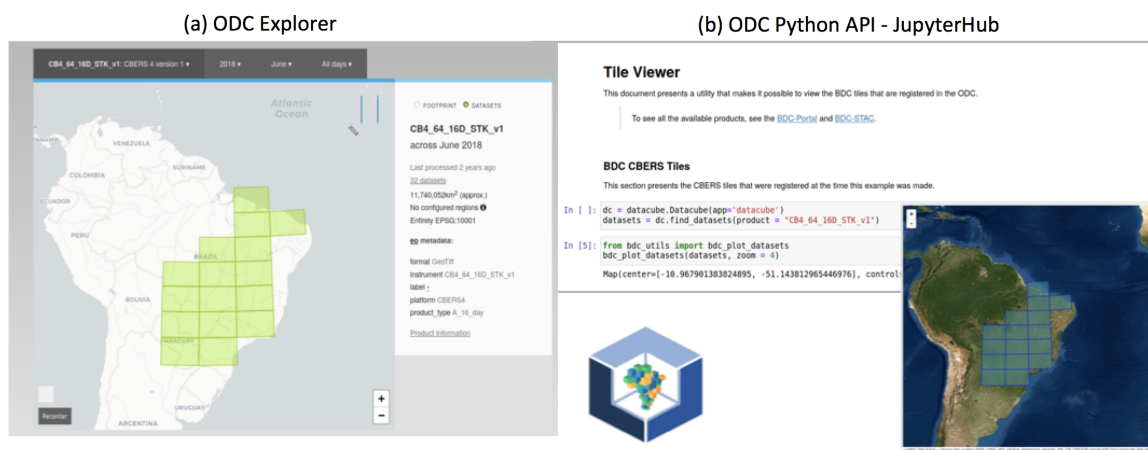


Figure 8. Accessing and processing CBERS-4 data cubes of the Brazilian Cerrado biome with the Open Data Cube framework: (a) ODC Explorer and (b) ODC Python API with JupyterHub.

The applications include the Satellite Image Time Series (sits) package, written in R language, which started to be developed in the e-Sensing project [32]. This package contains functions to retrieve time series from image collections and data cubes, different visualization methods for image time series, smoothing functions for noisy time series, and distinct clustering algorithms, including dendrograms and self-organizing maps (SOM). It also implements interfaces with other R packages that contain machine learning methods for time series classification, providing a unified environment for satellite image time series analysis. In the BDC project, we are improving the sits package with new methods and interfaces with the BDC data cubes and Web services. Forest Monitor is a Web application that uses the image collections and data cubes created by the BDC project to identify deforestation alerts. The DETER project team is using this application.

5. Land Use and Cover Mapping from EO Data Cubes

This section presents an annual land use and cover mapping method using the time series extracted from the data cubes created in the BDC project. Figure 9 describes the process of creating land use and cover maps from EO data cubes using satellite image time series analysis and deep learning methods.

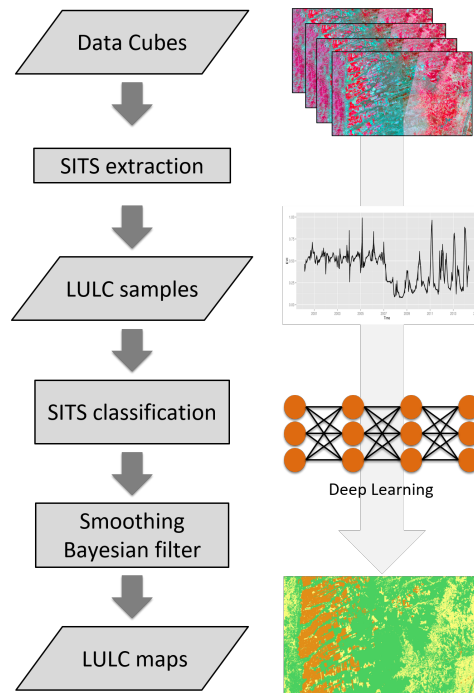


Figure 9. Process to create land use and land cover (LULC) maps from Earth observation (EO) data cubes using satellite image time series (SITS) analysis and deep learning methods.

5.1. Study Area and Data

For this mapping, we selected a study area located in the Bahia state, Brazil, between the Cerrado and Caatinga biomes, and used the monthly data cubes CBERS-4, Landsat-8, and Sentinel-2, based on stack compositing function as show in Figure 10. This region is known for the expansion of agriculture and livestock, which has been happening over the last few years in an intensive way [44].

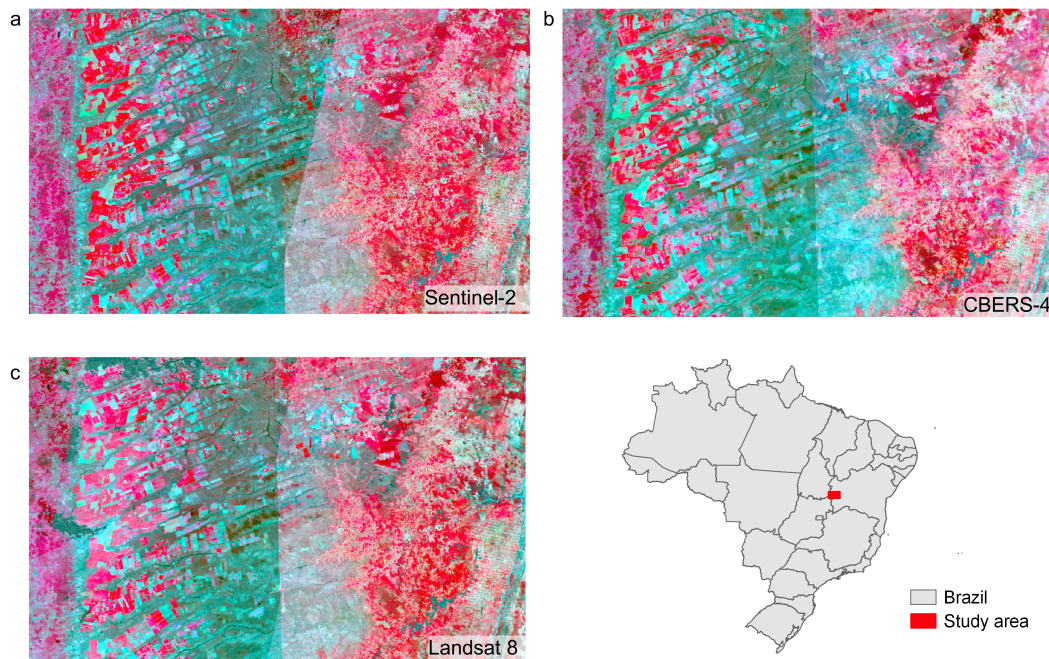


Figure 10. (a) Sentinel-2, (b) CBERS-4, and (c) Landsat-8 data cube images of the study area in January 2019.

In this experiment, we used time series of a agricultural calendar year, from September of 2018 to August of 2019, extracted from these three data cubes. For this study area, remote sensing specialists collected land use and cover samples using visual interpretation of high-resolution images and data cube time series. These data sets includes 922 samples of three classes: natural vegetation (422), pasture (258), and agriculture (242). Figure 11 shows the NDVI time series (September of 2018 to August of 2019) extracted from the CBERS-4 data cube of the these three classes.

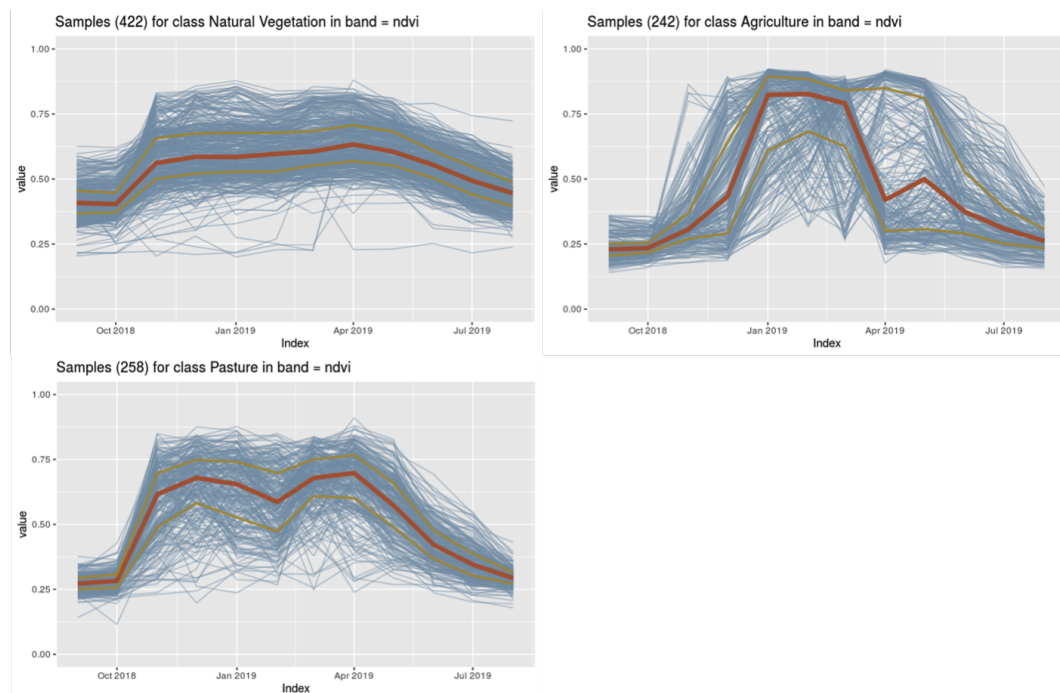


Figure 11. NDVI time series (September of 2018 to August of 2019) extracted from the CBERS-4 data cube of the three land use and cover classes: natural vegetation, pasture, and agriculture.

5.2. Classification Process

The classification was made using sits R package running in an AWS EC2 service and using the following data cube bands: red, green, blue, and near-infrared (NIR) along with the indices NDVI and EVI. We used these bands because they are common to the three data cubes. First, we collected the time series associated with the sample locations from each data cube and used them to train the classifier.

To classify the data cubes, we trained three different multi-layer perceptrons for a deep learning classification network [45], one for each data cube, using the the time series associated with sample locations. We used sits R package to train the models and classify the time series. The models consisted of five hidden layers with 512 neurons in a complete network architecture, and was trained through back propagation using adaptive learning optimization (Adam) and a rectilinear activation unit (ReLu) function.

The classification output is a map of probabilities for each pixel belonging to a land use and cover class. The majority of methods using this approach select the most probable class. Instead, we use the Bayesian smoothing method to select the pixel class, as proposed by Simoes et al. [11]. This method considers the pixel's mean and variance of the neighbors to compute the posterior Bayesian probabilities and then reanalyze the most probable class to each pixel.

5.3. Results

The classification process was applied to the three data cubes, generating three land use and cover maps shown in Figure 12.

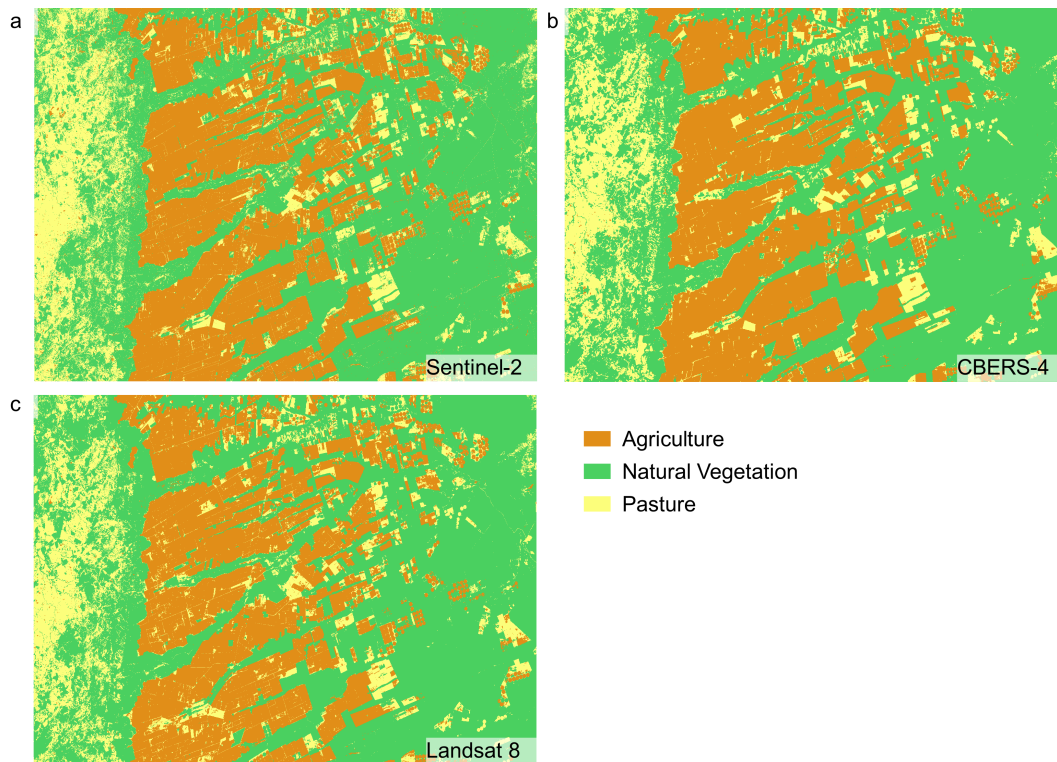


Figure 12. Deep learning classification using (a) Sentinel-2, (b) CBERS-4, and (c) Landsat 8 stack compositing monthly data cubes, from September 2018 to August 2019.

We validated the accuracy of the classified maps using the good practice guidelines by Olofsson et al. [46]. The validation was done independently for each map using the PRODES Cerrado data of 2019 [25]. This data consisted of anthropized areas in the Cerrado biome. To label the validation samples, we use the WLTS service.

We merged our agriculture and pasture classes to make them compatible with the anthropic class of PRODES Cerrado. Therefore, we obtained a classification with two classes natural vegetation and anthropic (pasture and Agriculture). We used these classes as strata to validate the maps. We assumed a target standard error of overall map accuracy was 0.01 [46] and the user accuracy (UA) of 0.8 for both natural vegetation and anthropic, which resulted in 1600 samples. We allocated an equal number of samples, 800 for each class. The classifications' accuracy are shown in Table 2.

Table 2. Producer's accuracy (PA), user's accuracy (UA), and overall accuracy (OA) of the CBERS-4, Sentinel-2, and Landsat 8 data cube classifications.

Accuracy	CBERS-4	Sentinel-2	Landsat 8
PA Anthropic	0.81	0.90	0.94
PA Nat. Veg.	0.67	0.84	0.85
UA Anthropic	0.71	0.85	0.86
UA Nat. Veg.	0.78	0.90	0.94
OA	0.74	0.87	0.90

Although the classifications are from the same location and use the same attributes, the validation results indicated that there are differences in the representations of the targets in each data cube. This suggests that each data cube should be classified with different machine learning algorithm parameters. Additionally, other methodological steps could be included to improve the performance of the classifications. For example, Santos et al. [30] proposes a method for the detection and selection of representative samples, excluding outliers or samples from different classes with similar time series

profiles from the data set. Moreover, Chaves et al. [47] suggests that the use of SWIR and Redge-Edge bands produces better land use and cover classification results. This indicates that there is room to improve further the results of Sentinel-2 and Landsat-8 data cubes.

6. Final Remarks and Future Directions

The first version (1.0) of the ARD image collections and data cubes produced by the BDC project is already available in the project Web portal and Web services for the Brazilian Cerrado biome. The image collections consist of surface reflectance data from CBERS-4/AWFI, Landsat-8/OLI for the entire biome ranging from 2017 to 2020, and specific tiles for Sentinel-2/MSI (A and B) considering the same period. From these image collections, the data cubes were generated for the same areas and periods using a 16 days time step and the stack temporal composing function.

To generate the ARD and data cubes, we developed a set of Python scripts and an application with a graphical interface called Data Cube Builder. These scripts and the application can be used to create ARD and data cubes for other regions, including other countries in South America. We have two versions of this application, one that runs on AWS using Lambda services and another running in our computer network at INPE. Therefore, users can run these scripts and applications to create data cubes based on their specific needs, such as using other time steps or temporal composition functions. All software products described in this paper are available in our GitHub area <https://github.com/brazil-data-cube>.

The ARD image collections, data cubes, and software products that we are producing and developing in the BDC project are useful for different applications that investigate the dynamics of an environment over time based on remote sensing images. Such applications include land use and cover change, deforestation, fire monitoring, and the urbanization process. In this paper, we present land use and cover mapping as an example of an application that benefits from the BDC data cubes and software products.

Other similar projects that generate ARD and data cubes for specific countries, such as Swiss Data Cube (SDC) [14], Digital Earth Australia (DE Australia) [12], and Digital Earth Africa (DE Africa) [18], are using the ODC framework. These other projects follow a similar approach for making data available, analysis, and processing. The data produced by these projects can be accessed without restrictions through visualization portals, Web catalogs, and OGC services. The SDC uses the Geoserver server, while the other two use ODC to provide OGC services. Both SDC and DE Australia do not offer public access and processing data via the ODC Python API. These resources are restricted to authorized researchers. The DE Africa, on the other hand, allows any registered user to have access to a cloud-based JupyterLab environment to perform their analysis and processing. This functionality is still provisional, and at the moment, offers few computational resources for each user. In the same direction, the BDC project provides a portal for viewing the data sets, a catalog for accessing the collections via the STAC protocol, and a JupyterHub environment for the project researchers. Besides that, we provide an interface with the ODC platform, as described in Section 4.2.

In this first version, each data cube is created from images of a specific sensor. However, we are investigating methods to create “harmonized” data cubes by mixing images from distinct sensors. Recently, efforts have been made to harmonize images of multiple multi-spectral sensors to enhance the number of Earth surface observations. A “harmonized” surface reflectance product is obtained when required steps (radiometric, spectral, geometric, spatial, and others) have been applied to create a consistent data set to time series analysis as though it came from a single sensor [48]. One of the most promising approaches has been the c-factor, since it can be used to generate Nadir BRDF adjusted reflectance (NBAR) data for Landsat [49] and Sentinel-2 [50]. Other data cube initiatives have also used this method, such as the Harmonized Landsat Sentinel (HLS) [51,52], to combine Landsat-8 and Sentinel-2 data. Based on these studies, the Brazil Data Cube project team is also evaluating the impact of atmospheric correction algorithms and started developing harmonization procedures to produce Data Cubes containing an image from both satellite sensors.

About future directions—we intend to create data cubes from synthetic aperture radar (SAR) data. SAR data have low atmospheric influence in the microwave spectrum range and thus are useful for environmental monitoring in tropical regions. Sentinel-1 satellite has generated free, open, regular, and standardized SAR images since 2014. These Sentinel-1 images have been used for land use and cover monitoring with good results [53,54]. We hope to improve the land use and cover mapping in the BDC project using SAR data cubes.

Author Contributions: Conceptualization, K.R.F., G.R.Q., L.V., G.C., and R.C.; methodology, K.R.F., G.R.Q., L.V., G.C., R.C., M.C.A.P., M.E.D.C., A.H.S., R.E.O.S., V.C.F.G., L.A.S., M.C.Z., R.F.B.M., T.S.K., A.R.S., and F.Z.; software, K.R.F., G.R.Q., L.V., J.S.A., C.A.N., R.C., R.E.O.S., R.F.B.M., R.W.C., V.C.F.G., L.A.S., M.C.Z., F.Z., and A.R.S.; validation, K.R.F., G.R.Q., L.V., G.C., R.C., J.G.F., M.C.A.P., M.E.D.C., A.H.S., R.E.O.S., V.C.F.G., L.A.S., M.C.Z., R.F.B.M., A.R.S., and F.Z.; formal analysis, G.C., R.C., J.G.F., A.H.S., R.E.O.S., R.F.B.M., and A.R.S.; investigation, K.R.F., G.R.Q., L.V., G.C., R.C., M.C.A.P., M.E.D.C., A.H.S., R.E.O.S., V.C.F.G., L.A.S., M.C.Z., R.F.B.M., T.S.K., A.R.S., and F.Z.; data curation, J.G.F.; writing—original draft preparation, K.R.F., G.R.Q., L.V., M.C.A.P., M.E.D.C., A.H.S., R.E.O.S., R.F.B.M., T.S.K., A.R.S., and J.G.F.; writing—review and editing, K.R.F., G.R.Q., and L.V.; visualization, C.A.N. and R.W.C.; project administration, K.R.F., G.R.Q., and L.V.; funding acquisition, L.M.G.F., K.R.F., G.R.Q., and L.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Amazon Fund through the financial collaboration of the Brazilian Development Bank (BNDES) and the Foundation for Science, Technology and Space Applications (FUNCATE), process 17.2.0536.1; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES), finance code 001; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); Group on Earth Observations (GEO) and Amazon Web Services (AWS) under the Earth Observation Cloud Credits Programme.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Foley, J.A.; DeFries, R.; Asner, G.P.; Barford, C.; Bonan, G.; Carpenter, S.R.; Chapin, F.S.; Coe, M.T.; Daily, G.C.; Gibbs, H.K.; et al. Global Consequences of Land Use. *Science* **2005**, *309*, 570–574. [[CrossRef](#)] [[PubMed](#)]
2. Gomez, C.; White, J.C.; Wulder, M.A. Optical remotely sensed time series data for land cover classification: A review. *J. Photogramm. Remote Sens.* **2016**, *116*, 55–72. [[CrossRef](#)]
3. Soille, P.; Burger, A.; Marchi, D.D.; Kempeneers, P.; Rodriguez, D.; Syrris, V.; Vasilev, V. A versatile data-intensive computing platform for information retrieval from big geospatial data. *Future Gener. Comput. Syst.* **2018**, *81*, 30–40. [[CrossRef](#)]
4. Gomes, V.C.; Queiroz, G.R.; Ferreira, K.R. An Overview of Platforms for Big Earth Observation Data Management and Analysis. *Remote Sens.* **2020**, *12*, 1253. [[CrossRef](#)]
5. Woodcock, C.E.; Loveland, T.R.; Herold, M.; Bauer, M.E. Transitioning from Change Detection to Monitoring with Remote Sensing: A Paradigm Shift. *Remote Sens. Environ.* **2020**, *238*, 111558. [[CrossRef](#)]
6. Maus, V.; Câmara, G.; Cartaxo, R.; Sanchez, A.; Ramos, F.M.; de Queiroz, G.R. A Time-Weighted Dynamic Time Warping Method for Land-Use and Land-Cover Mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3729–3739. [[CrossRef](#)]
7. Belgiu, M.; Csillik, O. Sentinel-2 Cropland Mapping Using Pixel-Based and Object-Based Time-Weighted Dynamic Time Warping Analysis. *Remote Sens. Environ.* **2018**, *204*, 509–523. [[CrossRef](#)]
8. Picoli, M.C.A.; Camara, G.; Sanches, I.; Simões, R.; Carvalho, A.; Maciel, A.; Coutinho, A.; Esquerdo, J.; Antunes, J.; Begotti, R.A.; et al. Big earth observation time series analysis for monitoring Brazilian agriculture. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 328–339. [[CrossRef](#)]
9. Kong, Y.L.; Huang, Q.; Wang, C.; Chen, J.; Chen, J.; He, D. Long short-term memory neural networks for online disturbance detection in satellite image time series. *Remote Sens.* **2018**, *10*, 452. [[CrossRef](#)]
10. Bendini, H.N.; Fonseca, L.M.G.; Schwieder, M.; Körting, T.S.; Rufin, P.; Sanches, I.D.A.; Leitão, P.J.; Hostert, P. Detailed agricultural land classification in the Brazilian cerrado based on phenological information from dense satellite image time series. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *82*, 101872. [[CrossRef](#)]
11. Simoes, R.; Picoli, M.C.; Camara, G.; Maciel, A.; Santos, L.; Andrade, P.R.; Sánchez, A.; Ferreira, K.; Carvalho, A. Land use and cover maps for Mato Grosso State in Brazil from 2001 to 2017. *Sci. Data* **2020**, *7*, 1–10. [[CrossRef](#)] [[PubMed](#)]

12. Lewis, A.; Oliver, S.; Lymburner, L.; Evans, B.; Wyborn, L.; Mueller, N.; Raevksi, G.; Hooke, J.; Woodcock, R.; Sixsmith, J.; et al. The Australian Geoscience Data Cube—Foundations and Lessons Learned. *Remote Sens. Environ.* **2017**, *202*, 276–292. [[CrossRef](#)]
13. Siqueira, A.; Tadono, T.; Rosenqvist, A.; Lacey, J.; Lewis, A.; Thankappan, M.; Szantoi, Z.; Goryl, P.; Labahn, S.; Ross, J.; et al. CEOS Analysis Ready Data For Land—An Overview on the Current and Future Work. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 5536–5537. [[CrossRef](#)]
14. Giuliani, G.; Chatenoux, B.; De Bono, A.; Rodila, D.; Richard, J.P.; Allenbach, K.; Dao, H.; Peduzzi, P. Building an Earth Observations Data Cube: Lessons Learned from the Swiss Data Cube (SDC) on Generating Analysis Ready Data (ARD). *Big Earth Data* **2017**, *1*, 100–117. [[CrossRef](#)]
15. Appel, M.; Pebesma, E. On-Demand Processing of Data Cubes from Satellite Image Collections with the Gdalcubes Library. *Data* **2019**, *4*, 92. [[CrossRef](#)]
16. Asmaryan, S.; Muradyan, V.; Tepanosyan, G.; Hovsepyan, A.; Saghatlyan, A.; Astsatryan, H.; Grigoryan, H.; Abrahamyan, R.; Guigoz, Y.; Giuliani, G. Paving the Way towards an Armenian Data Cube. *Data* **2019**, *4*, 117. [[CrossRef](#)]
17. Maso, J.; Zabala, A.; Serral, I.; Pons, X. A Portal Offering Standard Visualization and Analysis on top of an Open Data Cube for Sub-National Regions: The Catalan Data Cube Example. *Data* **2019**, *4*, 96. [[CrossRef](#)]
18. Killough, B. The Impact of Analysis Ready Data in the Africa Regional Data Cube. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 5646–5649. [[CrossRef](#)]
19. Camara, G.; Assis, L.F.; Ribeiro, G.; Ferreira, K.R.; Llapa, E.; Vinhas, L. Big Earth Observation Data Analytics: Matching Requirements to System Architectures. In Proceedings of the 5th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data, BigSpatial '16, San Francisco, CA, USA, 31 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 1–6. [[CrossRef](#)]
20. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**. [[CrossRef](#)]
21. Giuliani, G.; Camara, G.; Killough, B.; Minchin, S. Earth Observation Open Science: Enhancing Reproducible Science Using Data Cubes. *Data* **2019**, *4*, 147. [[CrossRef](#)]
22. Ponzoni, F.J.; Junior, J.Z.; Lamparelli, R.A. In-flight absolute calibration of the CBERS-2 CCD sensor data. *Acad. Bras. Ciências* **2008**, *80*, 373–380. [[CrossRef](#)]
23. Diniz, C.G.; Souza, A.A.; Santos, D.C.; Dias, M.C.; Luz, N.C.; Moraes, D.R.V.; Maia, J.S.; Gomes, A.R.; Narvaes, I.; Valeriano, D.M.; et al. DETER-B: The New Amazon Near Real-Time Deforestation Detection System. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3619–3628. [[CrossRef](#)]
24. INPE. PRODES Project: Brazilian Amazon Forest Monitoring by Satellite. Available online: <http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes> (accessed on 28 December 2019).
25. FG Assis, L.F.; Ferreira, K.R.; Vinhas, L.; Maurano, L.; Almeida, C.; Carvalho, A.; Rodrigues, J.; Maciel, A.; Camargo, C. TerraBrasilis: A Spatial Data Analytics Infrastructure for Large-Scale Thematic Mapping. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 513. [[CrossRef](#)]
26. Almeida, C.A.D.; Coutinho, A.C.; Esquerdo, J.C.D.M.; Adami, M.; Venturieri, A.; Diniz, C.G.; Dessay, N.; Durieux, L.; Gomes, A.R. High spatial resolution land use and land cover mapping of the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data. *Acta Amaz.* **2016**, *46*, 291–302. [[CrossRef](#)]
27. Kintisch, E. Improved monitoring of rainforests helps pierce haze of deforestation. *Science* **2007**, *316*, 536–537. [[CrossRef](#)] [[PubMed](#)]
28. Maus, V.; Câmara, G.; Appel, M.; Pebesma, E. dtwSat: Time-Weighted Dynamic Time Warping for Satellite Image Time Series Analysis in R. *J. Stat. Softw. Artic.* **2019**, *88*, 1–31. [[CrossRef](#)]
29. Xi, W.; Du, S.; Wang, Y.C.; Zhang, X. A Spatiotemporal Cube Model for Analyzing Satellite Image Time Series: Application to Land-Cover Mapping and Change Detection. *Remote Sens. Environ.* **2019**. [[CrossRef](#)]
30. Santos, L.; Ferreira, K.R.; Picoli, M.; Camara, G. Self-Organizing Maps in Earth Observation Data Cubes Analysis. In *Advances in Self-Organizing Maps, Learning Vector Quantization, Clustering and Data Visualization*; Vellido, A., Gibert, K., Angulo, C., Martín Guerrero, J.D., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 70–79.

31. Lu, M.; Pebesma, E.; Sanchez, A.; Verbesselt, J. Spatio-temporal change detection from multidimensional arrays: Detecting deforestation from MODIS time series. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 227–236. [[CrossRef](#)]
32. Camara, G.; Simoes, R.; Andrade, P.R.; Maus, V.; Sánchez, A.; de Assis, L.F.F.G.; Santos, L.A.; Ywata, A.C.; Maciel, A.M.; Vinhas, L.; et al. *e-sensing/sits: Version 1.12.5*; Zenodo: Geneva, Switzerland, 2018. [[CrossRef](#)]
33. Vinhas, L.; de Queiroz, G.R.; Ferreira, K.R.; Camara, G. Web services for big earth observation data. *Rev. Bras. Cartogr.* **2017**, *69*, 5.
34. Müller, M.; Bernard, L.; Brauner, J. Moving code in spatial data infrastructures - Web service based deployment of geoprocessing algorithms. *Trans. GIS* **2010**, *14*, 101–118. [[CrossRef](#)]
35. Amazon Web Services. Open Data on AWS. Available online: <https://aws.amazon.com/opendata/> (accessed on 26 March 2020).
36. Ferreira, K.R.; Queiroz, G.R.; Camara, G.; Souza, R.C.M.; Vinhas, L.; Marujo, R.F.B.; Simoes, R.E.O.; Noronha, C.A.F.; Costa, R.W.; Arcanjo, J.S.; et al. Using Remote Sensing Images and Cloud Services on AWS to Improve Land Use and Cover Monitoring. In Proceedings of the 2020 IEEE Latin American GRSS ISPRS Remote Sensing Conference (LAGIRS), Santiago, Chile, 22–26 March, 2020; pp. 558–562.
37. Vermote, E.; Justice, C.; Claverie, M.; Franch, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* **2016**, *185*, 46–56. [[CrossRef](#)]
38. Da Silva, M.A.O.; de Andrade, A.C. Geração de imagens de reflectância de um ponto de vista geométrico. *Rev. Bras. Geomat.* **2013**, *1*, 23. [[CrossRef](#)]
39. Qiu, S.; Zhu, Z.; He, B. Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. *Remote Sens. Environ.* **2019**, *231*, 111205. [[CrossRef](#)]
40. Sanchez, A.H.; Picoli, M.C.A.; Camara, G.; Andrade, P.R.; Chaves, M.E.D.; Lechler, S.; Soares, A.R.; Marujo, R.F.B.; Simões, R.E.O.; Ferreira, K.R.; et al. Comparison of Cloud Cover Detection Algorithms on Sentinel-2 Images of the Amazon Tropical Forest. *Remote Sens.* **2020**, *12*, 1284. [[CrossRef](#)]
41. Sattler, K.U. Data Quality Dimensions. *Encycl. Database Syst.* **2016**, 1–5. [108-2](#). [[CrossRef](#)]
42. STAC. SpatioTemporal Asset Catalogs. Available online: <https://stacspec.org/> (accessed on 2 October 2020).
43. OGC. OGC Standards and Supporting Documents. Available online: <http://www.opengeospatial.org/standards/> (accessed on 1 October 2020).
44. De Oliveira, S.N.; de Carvalho Júnior, O.A.; Gomes, R.A.T.; Guimarães, R.F.; McManus, C.M. Landscape-fragmentation change due to recent agricultural expansion in the Brazilian Savanna, Western Bahia, Brazil. *Reg. Environ. Chang.* **2017**, *17*, 411–423. [[CrossRef](#)]
45. Hastie, T.; Tibshirani, R. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*; Springer: New York, NY, USA, 2009.
46. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good Practices for Estimating Area and Assessing Accuracy of Land Change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [[CrossRef](#)]
47. Chaves, M.E.D.; Picoli, M.C.A.; Sanches, I.D. Recent Applications of Landsat 8/OLI and Sentinel-2/MSI for Land Use and Land Cover Mapping: A Systematic Review. *Remote Sens.* **2020**, *12*, 3062. [[CrossRef](#)]
48. Claverie, M.; Masek, J.G.; Ju, J.; Dungan, J.L. *Harmonized Landsat-8 Sentinel-2 (HLS) Product User's Guide: Product Version 1.3*; National Aeronautics and Space Administration (NASA): Washington, DC, USA, 2017.
49. Roy, D.P.; Zhang, H.K.; Ju, J.; Gomez-Dans, J.L.; Lewis, P.E.; Schaaf, C.B.; Sun, Q.; Li, J.; Huang, H.; Kovalsky, V. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sens. Environ.* **2016**, *176*, 255–271. [[CrossRef](#)]
50. Roy, D.P.; Li, J.; Zhang, H.K.; Yan, L.; Huang, H.; Li, Z. Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance. *Remote Sens. Environ.* **2017**, *199*, 25–38. [[CrossRef](#)]
51. Claverie, M.; Ju, J.; Masek, J.G.; Dungan, J.L.; Vermote, E.F.; Roger, J.C.; Skakun, S.V.; Justice, C. The Harmonized Landsat and Sentinel-2 Surface Reflectance Data Set. *Remote Sens. Environ.* **2018**, *219*, 145–161. [[CrossRef](#)]
52. Franch, B.; Vermote, E.; Skakun, S.; Roger, J.c.; Masek, J.; Ju, J.; Villaescusa-nadal, J.L.; Santamaria-artigas, A. A Method for Landsat and Sentinel 2 (HLS) BRDF Normalization. *Remote Sens.* **2019**, *2*, 632. [[CrossRef](#)]
53. Steinhausen, M.J.; Wagner, P.D.; Narasimhan, B.; Waske, B. Combining Sentinel-1 and Sentinel-2 data for improved land use and land cover mapping of monsoon regions. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 595–604. [[CrossRef](#)]

54. Nicolau, A.P.; Flores-Anderson, A.; Griffin, R.; Herndon, K.; Meyer, F.J. Assessing SAR C-band data to effectively distinguish modified land uses in a heavily disturbed Amazon forest. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *94*, 102214. [[CrossRef](#)]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).