FAPESP Research Program in e-science - Grant 2014/08398-6

Project duration: January 01, 2015 – 31 December 2018

e-Sensing:

Big Earth observation data analytics

for land use and land cover change information

First yearly report: 01 January 2016 – 31 December 2016

Host institution: INPE – Instituto Nacional de Pesquisas Espaciais

Principal investigator: Gilberto Câmara (INPE)

Research team: Ieda Sanches, Karine Ferreira, Leila Fonseca, Lubia Vinhas, Isabel Escada, Gilberto Queiroz, Luiz Maurano, João Vianei Soares, Julio d'Alge, Pedro Andrade, Ricardo Cartaxo, Thales Körting.

Additional investigators: Eduardo Llapa (INPE)

Project website: http://www.esensing.org

Table of Contents

1 Overview of the project objectives

This document provides the second year report of the "e-sensing" FAPESP project (grant 2014/08398-6), and describes the activities carried out during the period 01.01.2016 to 31.12.2016. We will use numbers (such as [10]) to refer to the list of papers published by us in 2016, available in the References section.

Currently, scientists ignore the time reference inherent to Earth observation data, producing land cover maps taking either a single or at most two time references. As a result, only a small part of the big data sets produced by remote sensing satellite are ever used. This leads to an important research question: *How can we use e-science methods and techniques to substantially improve the extraction of land use and land cover change information from big Earth Observation data sets in an open and reproducible way?*

In response to this challenge, *our project will conceive, build and deploy a new type of knowledge platform for organization, access, processing and analysis of big Earth observation data*. The key elements of this knowledge platform are:

1. A scientific database based on the SciDB innovative array database management system, capable of managing large remote sensing data sets.
2. An innovative set of spatiotemporal image analysis methods, mostly based in analysis of satellite image time series. These methods are all developed as open source software to promote reproducibility.

The innovative infrastructure developed in the project will be used for new types of information extraction from Earth observation data, focused on land cover and land use change of large data sets. Our knowledge platform will allow scientists to perform data analysis directly on big data servers. Scientists will be then able to develop completely new algorithms that can seamlessly span partitions in space, time, and spectral dimensions.

We aim to make two important contributions:

1. New database methods and techniques that use array databases to build a geographical information system that handles big spatial data.
2. New data analysis, data mining, and image processing methods to extract land change information from large Earth observation data sets.

2 Main results of year 2 (January – December 2016)

During 2016, our most relevant results were:

1. Implementation an operational software infrastructure for handling and processing of large Earth observation data sets [12][27].
2. Evaluation and comparison of big data architectures between array databases (SciDB) and MapReduce methods (Hadoop) [20].
3. Development of spatio-temporal extensions to PostGIS [23] [26], for handling spatio-temporal data from different data sources [11], and for extracting metadata from big Earth observation data [17].
4. Completion of the development of a TWDTW, new method for satellite image time series analysis using a time-weighted version of the dynamical time warping data mining technique [3].
5. Evaluation of the use of the BFAST break-detection algorithms for detection of deforestation using remote sensing time series [2].
6. The development of techniques for data mining of trajectories of forest degradation in Amazonia [5][6].
7. The proposal of an event-based ontological model for extracting information from big Earth observation data sets [16].
8. Validation and testing of baseline methods for automated classification of agricultural areas [1][11][13][22][25].

Overall, the project is progressing as expected. During the second year, the team has made significant progress in the infrastructure and the analysis methods. These results will allow important new results on land user and land cover change classifications to be achieved in years 3 and 4.

The two post-docs associated to the development of applications of big data analytics, respectively in Forestry and Agriculture, have been selected in 2016. After a period of training to learn how to use the methods developed by the project team, they are now ready to make important contributions. Thus, the project is on track to achieve its stated aims.

3 Detailed description of the results in Year 2 (2016)

This section describes the results of the project in 2016. In the presentation, we follow the project organization in three work packages (WP), and associated milestones, as laid out in the proposal:

1. *WP 1 – Databases*: research and development associated with using array databases to store large Earth observation data sets and developing workflows and methods for efficient storage, access and processing of large data, reproducibly.

2. *WP 2 – Data analysis*: R&D on spatiotemporal techniques for extracting change information on large Earth observation data sets, relevant for forestry applications; include novel time series applications for remote sensing data, and combined time series and multi-temporal image processing.

3. *WP 3 – Use case development*: case studies of forestry and agriculture applications that use large Earth observation data sets. These use cases will validate the methods and data developed by the other work packages.

To help the review of this report, we first present the table of milestones presented in the project proposal. We will then consider each the proposed milestones, stating whether it has been fulfilled or delayed.

For each milestone, we preview the result more directly associated with it. The rest of the results of the project can be found in the References section. All of the papers published by members of the research teams that are associated to the project are available at the project's website: http://www.esensing.org.

TABLE 1

PLANNED MILESTONES: RESULTS AFTER MONTH 24

Green background	Target was met
Yellow background	Target was partially met
Red background	Target was delayed
Blue background	New task
White background	For later years

TASK	Month 12	Month 24	Month 36	Month 48
T1.1 **Building big EO databases**	M1.1.1. V1 of the database for use cases in Brazil	M1.1.2 V2 of database for use cases in Brazil	M1.1.3 V3 of database for use cases in Brazil	M 1.1.4 V4 of database for global use cases
T1.2 **Extend SciDB for geographical data handling**	M1.2.1 Integration of TerraLib and SciDB	M1.2.2 Algorithms for SciDB server-side processing	M1.2.3 Web service for SciDB server-side processing	M1.2.4 Extension of SciDB as a spatial data manager
T2.1 **Exploratory big data analysis**		M2.1.2 Interactive environment for data exploration	M2.1.3 Interactive environment for collaborative analysis	M2.1.4 Large-scale sharing with other research teams
T2.2 **Data analysis for big EO data**	M2.2.1 R-Big-EO time series analysis software (V1)	M2.2.2 R-Big-EO time series analysis software (V2)	M2.2.3 R-Big-EO space-time analysis software (V1)	M2.2.4 Big-EO space-time analysis software (V2)
T3.1 **Tropical forest change**	M3.1.1 Identification and selection of areas	M3.1.2 Preliminary detection of clear cut and degradation	M3.1.3 Detection of clear cut and degradation: final results	M3.1.4 Assessment of forest change alert methods
T3.2 **Tropical agriculture mapping**	M3.2.1 Identification and selection of areas	M3.2.2 Mapping of soybeans, maize and sugarcane	M3.2.3 Mapping of soybeans, maize, rice sugarcane, and wheat	M3.2.4 Assessment of agricultural mapping methods

3.1 Progress report on WP 1 - Big Earth observation databases

3.1.1 Task 1.1 - Building and deployment of big Earth observation databases to support data analysis and use cases

This task builds databases to be used by the project. In the proposal, we set the following milestone for month 24:

Milestone M1.1.2 – V2 of database for use cases in Brazil

The target of this milestone was to build a large remote sensing database containing the data needed for the use cases in Brazil in year 3. <u>This result has been achieved</u>.

As we described in the Year 1 report, we have chosen SciDB, an open source array database optimized for management of big data, that is suited for time series analysis of remote sensing imagery. When we arrange many images from the same areas as 3D arrays, each pixel is associated satellite image time series (Fig. 1a). Satellite image time series are important for land cover change analysis. Figure 1b shows an example of a time series of a vegetation index for a location in the Brazilian Amazonia from 2000 to 2008.

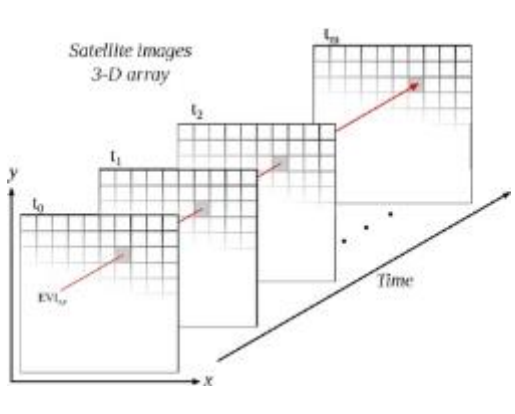

Figure 1 (a) – Organizing 2D images as 3D arrays (source: [3])

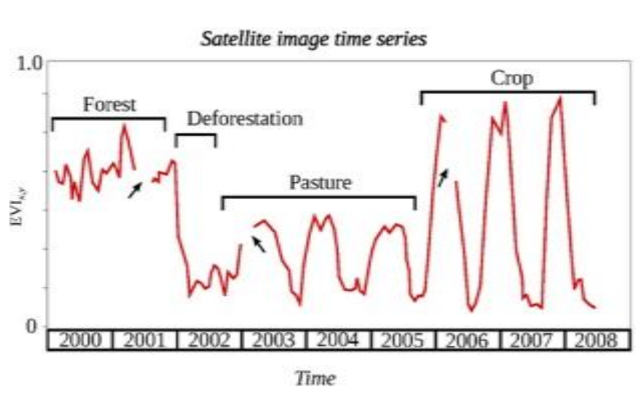

Figure 1 (b) – Example of a satellite image time series (MODIS EVI data) (source: [3]).

To organize image data as 3D arrays using SciDB, we have bought a server configuration, with support from FAPESP, consisting of:

1. 5 PowerEdge R730 servers, each CPU with 2 processors with 6 cores each.
2. 96 GB of main memory per server.
3. 13 TB of disk storage per server, with a total of 65 TB of storage.

On year 2, we have loaded in the servers the following data sets:

1. MODIS MOD09Q1 images at 250-meter resolution from 2000 to 2015 for the whole of South America, with 13,800 images associated to 3.11 x 1011 (317 billion) different satellite image time series.
2. Selected LANDSAT images at 30-meter resolution from 2000 to 2015 for selected areas of Amazonia, with 202 images associated to 200 million different satellite image time series.

3.1.2 Task 1.2 – Extend SciDB for geographical data handling

The purpose of this task was to use the array database SciDB for Earth observation applications. To do this, we need to develop methods that would allow us to process satellite image time series in SciDB.

Originally, we envisaged using the TerraLib software library, developed by INPE, as a source of data types and algorithms for geographical data. Our plan was to include an interface for SciDB in TerraLib. However, we have found out that this integration was not required to achieve our aims. This simplifies the design of the infra-structure and allows for easier reproducibility. Therefore, we decided to use web services as the primary interface for our big data analytical methods [26].

In this task, the project proposal has the following milestone for month 24:

Milestone M1.2.2 – Algorithms for SciDB server-side processing

The aim of this new milestone was to develop a series of methods for processing big EO data on array databases. *This target has been achieved*. The strategy chosen is described in the ACM GIS paper “Big earth observation data analytics: matching algorithms with system architectures” [12] and is resumed below.

When designing an architecture for big EO data, one needs to consider the needs of data analytics. One crucial observation is that researchers are most productive when working on familiar computing environments. Scientists like to test new ideas on small and well-known data sets. Only after they are satisfied with the experiments, they move up to work with big data. Therefore, an architecture for big Earth observation data analytics should meet important needs of the research community:

1. *Analytical scaling:* provide support for the full cycle of research, allowing algorithms developed at the desktop to run on big databases with minor changes.
2. *Software reuse*: allow researchers to adapt existing methods for big data with minimal reworking.
3. *Collaborative work*: enable results to be shared with the scientific community.
4. *Replication*: encourage research teams to build their own infrastructure.

Data scientists prefer to work on a simple software kernel where they can add new packages that encapsulate new analytical methods. For this reason, we have chosen the R suite of statistical tools as the environment to develop our analytical methods. R is the *lingua franca* of data analytics. Using R, researchers can scale up their methods, reuse previous work, and collaborate with their peers Thus, we have developed methods that allow scientists to execute their algorithms directly in big data servers (see Figure 2).

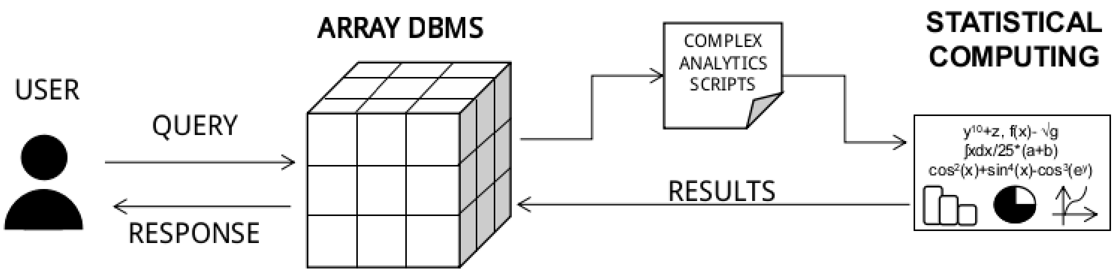

Figure 2 – Interface for server-side processing of analytical methods. Source:[12].

For efficient server-side processing, we optimized the organization of the array database SciDB for time series data analysis. Our cluster has a coordinator instance of SciDB and processing instances. Each processing instances of SciDB is assigned to data chunks that are independent. This organization takes advantage of the distributed CPUs, memory and disks to maximize parallel performance [11].

We also compared the performance of the array database SciDB with Hadoop, a more well-known MapReduce environment [12][20][27]. We found out that more work is required to adapt Hadoop to run R programs for EO data analysis than in the case of SciDB. The map/reduce model used by Hadoop requires preprocessing the image data into text files (or sequence files), while the array database represents image data directly. Thus, organizing image data for processing in Hadoop is more complicated and time-consuming than handling images in SciDB. In terms of performance, we obtain a similar result when running the TWDTW method in both Hadoop and SciDB solutions [12][20][27].

3.2 Progress report on WP 2 - Data analysis for big Earth observation data

3.2.1 Task 2.1 – Web-based exploratory big data analysis

In our original proposal, we envisaged making an integration between SciDB, TerraLib and the R software. However, during the development of the project we found out that this interface would not be required to achieve our goals of building an efficient set of software for space-time analysis. Instead, we have developed a web-based interface exploratory big data analysis. The exploratory data analysis interface is shown in Figures 3-5.

Figure 3 – Web-based exploratory big EO data analysis interface. In the left, there is a window allowing the user to select a training sample for later use in classification. In the right, the result of a classification of a point using the TWDTW algorithm is shown (source: e-sensing team).

Figure 4 – The figure shows the interface to gather samples in more detail. The user has selected locations in the image in the right and the interface retrieves the time series, using the WTSS service [26].

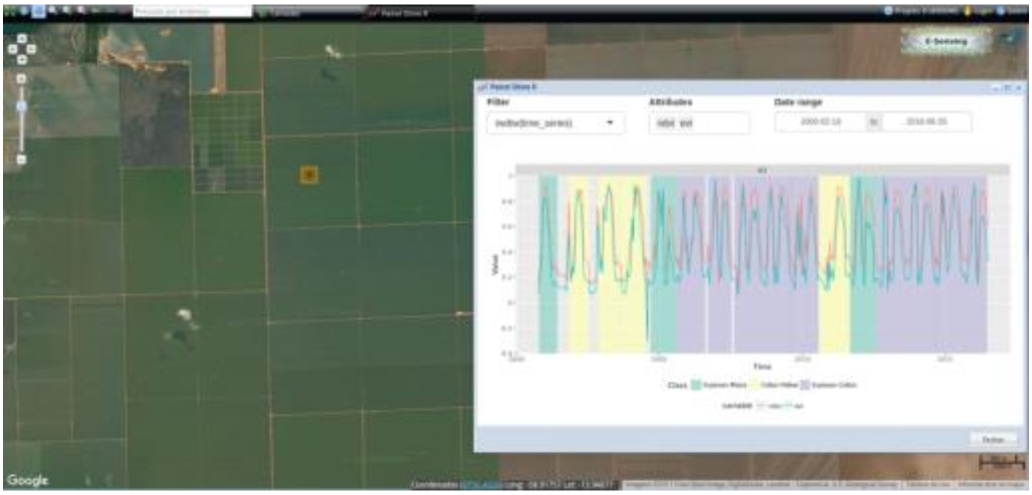

Figure 5 – Example of classification of one crop-producing location using the TWDTW data mining tools (source: e-sensing team).

The interfaces use the Web Time Series Service (WTSS, described in section 3.2.2) to obtain time series data. The user can then analyse the data using data mining algorithms such as TWDTW [3] and BFAST[1]. The user can also select training samples that can be later use to define patterns for remote sensing data mining. It works with the other R packages developed and under development by the project team: *WTSS, dtwSat, sits and WTSCS*, which are described below.

---

[1] Verbesselt, J., Hyndman , R ., Newnham , G ., and Culvenor, D. (2010). Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment, 114(1):106*

3.2.2 Task 2.2 - Space-time analysis of big Earth observation data for land change monitoring

This task aims to develop new methods for space-time analysis of big Earth observation data. It is expected to produce new research results, since it will be the first time that EO scientists have full access to large data sets to validate their data analysis methods.

In this task, the project proposal set the following milestone for month 24:

Milestone M2.2.2: Version 2 of R-Big-EO time series analysis software

This milestone has been achieved. Version 2 of the *R-big-EO time series analysis software*[2] has been designed and developed, consisting at present of the following R packages and additional software:

a) *WTSS* – web time series service, a lightweight service that interfaces with big EO data archives.
b) *WTSCS* – web time series classification service, a service that allows the user to request the server-side processing of large Earth observation data sets.
c) *Web-based exploratory data analysis* – the interface that allows the user to explore the big data sets, select training samples, perform analysis, and request server-side processing (described in section 3.2.1 above).
d) *sits* – an R package that provides the "glue" between the different services (WTSS, WTSCS) and classification and detection methods such as *dtwSat*.
e) *dtwSat* - R package that contains a new method for classification of satellite image time series: *time-weighted dynamic time warping* [3].

The WTSS is a lightweight web service. WTSS saves time when dealing with huge volumes of data because it has a flexible and simple API. It uses the JSON format instead of XML to deliver complex responses, which are then easily consumed by R [27]. The WTSS is fully operational and is used by the web-based interface (see above) and the other packages of the R-big-EO analysis software.

[2] In the original proposal, we called this software "big EO time series R package". We have changed the name to "R-Big-EO time series analysis software" because it includes different components that work together.

The WTSCS is a web service, currently under development, that will allow the user to request large-scale processing in a big data cluster. The WTSCS service goes beyond what is currently offered by OGC standards such as WPS and WCPS. Researchers will be able to develop and share new methods, working on a familiar R environment. Such a service that would allow significant progress on big EO data analytics [26]. The development of this service will meet the needs of project deliverable M1.2.3 (see Table 1 above).

The *sits* R package allows non-experts to perform complex data analysis of *satellite image time series* (hence "sits"). Users easily retrieve data from the server, and then use algorithms such as TWDTW [3] for processing. The package also provides interfaces to R analysis packages for analysis of seasonality[3], time series clustering[4] and time series smoothing. In Year 3 of the project, the *sits* package will include methods for frequency analysis, and neural network classification.

In year 1 of the project, we developed the *dtwSat* package, an R package that contains a new method for classification of satellite image time series: *time-weighted dynamic time warping*. In year 2, we improved the *dtwSat* package by including methods for classification validation and accuracy analysis. A paper describing this method was published in the *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (impact factor 2.7) [3]. The *dtwSat* package is also available at the R CRAN repository at the website cran.r-project.org/web/packages/dtwSat. In [3] we compared our classification with the MODIS land cover product, currently the only global land cover product available yearly (Table 2). Results show a better performance of TWDTW.

TABLE 2

Assessment of MODIS Land Cover and TWDTW for Porto dos Gaúchos (source:[3])

	Reliability		Accuracy	
Class	MODIS	TWDTW	MODIS	TWDTW
Forest	87%	94%	77%	88%
Pastureland	67%	88%	85%	85%
Cropland	89%	92%	75%	96%

[3] Verbesselt, J., Hyndman , R ., Newnham , G ., and Culvenor, D. (2010). Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment, 114(1):106*

[4] Alex Sarda, Comparing Time-Series Clustering Algorithms in R Using the dtwclust Package. Available at the CRAN R repository.

In the above table, the column *reliability* ("user's accuracy") shows the probability that a pixel labeled as a certain land-cover class in the map is really this class. The figures in column *accuracy* (also known as "producer's accuracy") refer to the probability that a certain land-cover of an area on the ground is classified as such.

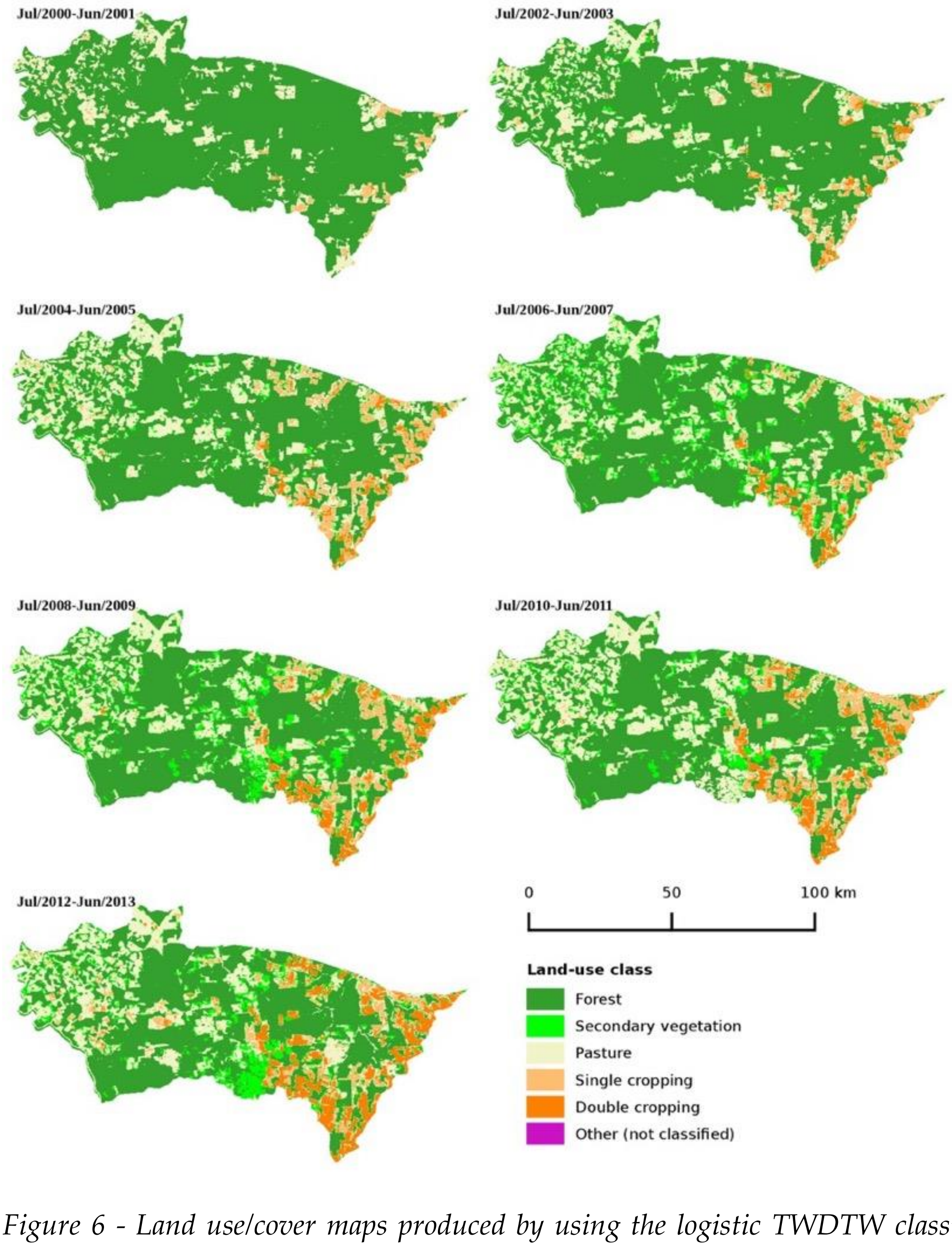

Figure 6 - Land use/cover maps produced by using the logistic TWDTW classification. Each map shows the classification for an agricultural year (from July to June) in Porto dos Gaúchos (source: [3]).

3.3 Progress report on WP 3 - Use case development

3.3.1 Task 3.1 - Specification and validation of tropical forest change alert methods and data

Our use cases address two important problems: (a) how to use time series and spatiotemporal analysis to support the DEGRAD system of monitoring forest degradation; (b) how to use time series data to support the DETER system for daily alerts of forest change. Our aim is not to replace INPE's existing systems, but to explore how automated methods can complement and enhance them. This work is being led by Prof. Dr. Isabel Escada [4][5][6].

In years 1 and 2, we did a lot of work on forest degradation, which is the most pressing and tougher scientific challenge. In this task, the project proposal has the following milestone for month 24:

Milestone M3.2.1: Preliminary detection of clear cut and degradation

This milestone has been partially accomplished. In this task, we have developed a method for trajectory data mining in forest degradation studies. Forest degradation is defined as defined as *the long-term and gradual reduction of canopy cover due to forest fire and unsustainable logging.* Typically, a mature forest is degraded when part of its tree cover is removed. In the Brazilian Amazonia, degradation is caused by either unsustainable logging practices or by forest fires (intentional or uncontrolled).

Understanding and identifying forest degradation is important, because it causes carbon emissions that have to be accounted for and leads to a loss of biodiversity caused by removal of important species. Also, degradation often (but not always) is associated to later actions that cause the full removal of forest cover (deforestation).

In Year 2, we carried out a study to define different types of land cover change trajectories [5][6]. In this study, the following trajectories have been identified:

a) *Converge*: a set of trajectories in a neighborhood, where all land cover classifications converge to the same final state. This pattern characterizes a convergence to a given state.
b) *Meet*: a set of trajectories in a given neighborhood that have a simultaneous transition to a specific final state.

c) *Flock*: a set of trajectories, which have the same sequence of states s initiated at time t. In this case, we try to identify the trajectories that have simultaneous state transitions over time, started at the same instant.

d) *Leadership*: a set of trajectories, which have the same sequence of states in a time window. The sequence that occurs first determines the leading trajectories.

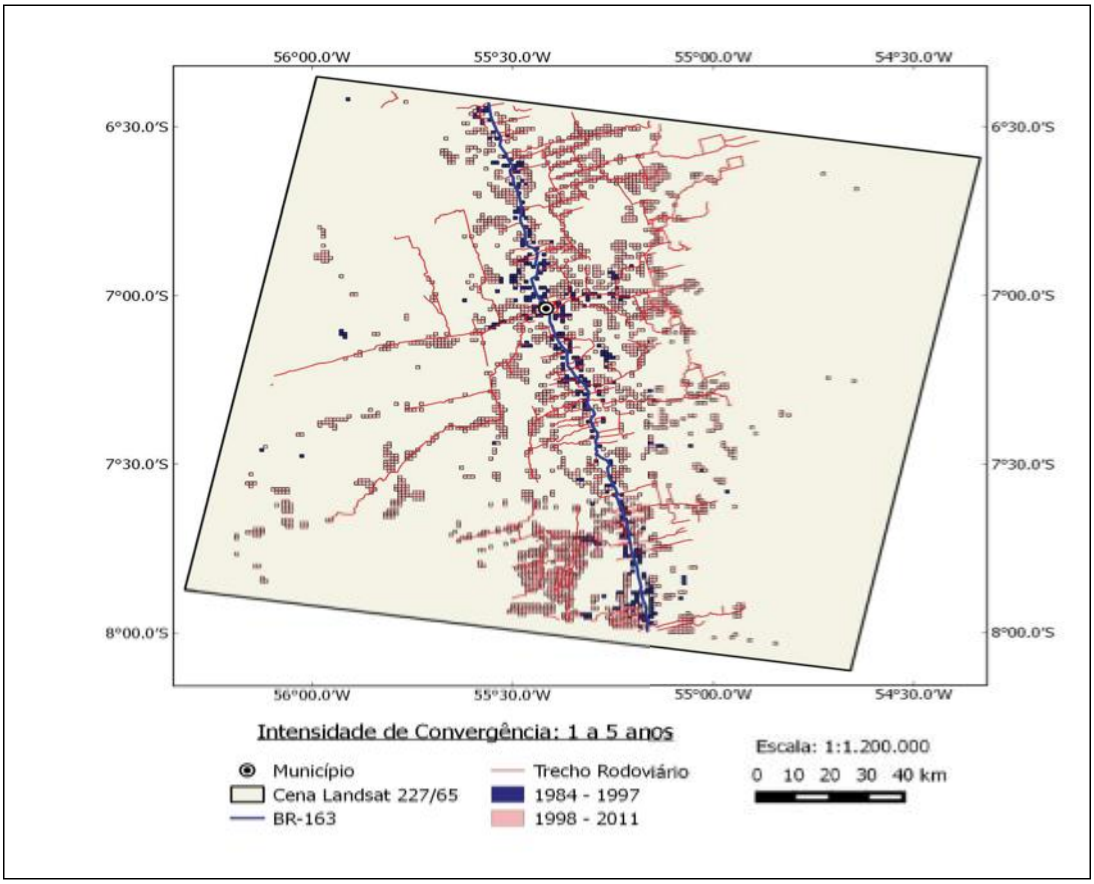

Figure 7 – Trajectories of high convergence in Novo Progresso, PA. In the highlighted areas, all land cover classifications converge to the same final state in 5 years or less. Source: [6].

This task suffered from a delay in selecting and hiring the post-doc specialist that will carry out the main actions required for producing the aims in Task 3.1. The post-doc (Rodrigo Begotti) has a PhD in Forest Resources from Escola Superior de Agricultura Luiz de Queiroz (ESALQ), USP and has started working in the project in the october of 2016. We expect that Rodrigo will make significant progress in 2017 and will be able to complete the deliverables of Task 3.1.

3.3.2 Task 3.2 - Specification and Validation of Tropical Agriculture Monitoring Methods and Data

The second use case is to apply satellite image time series analysis to improve the mapping and monitoring of Brazilian major agricultural commodities: soy, maize, sugarcane, rice and wheat. We will build on experiences by INPE's researchers, that have done extensive fieldwork and collected much ground truth data. These data sets will be used to validate our automated methods that work with big EO data. This work is led by Prof. Dr. Ieda Sanches [1][8][13][14][21].

In this task, the project proposal has the following milestone for month 24:

Milestone M3.2.2: Mapping of soybeans, maize and sugarcane

This milestone has been partially accomplished. The main drawback was the delay in selecting the post-doc to work on agricultural applications. After a careful selection process, we selected Dr. Michelle Picoli, who has a PhD in Agricultural Engeneering from Campinas State University (UNICAMP) and has also worked in CTBE (Bioethanol Technology Centre). Dr. Picoli only joined the project at the start of 2017.

As a preparatory work for the mapping of soybean, maize and sugarcane, the project team did an extensive work of data collection, so that we could obtain good training samples to classify agricultural areas. We obtained a data set that was used on an earlier paper on agricultural patterns done by a group from EMBRAPA[5]. Based on their data and on sample data collected by fieldwork by Prof. Dr. Ieda Sanches, we have been able to produce a detailed set of patterns associated to major crops in Brazil (see Figure 8). The preliminary results of using these patterns have shown the capacity of detecting different types of agricultural crops (see Figure 9) using the TWDTW algorithm [3].

[5] Damien Arvor, Margareth Meirelles, Vincent Dubreuil, Agnes Begue, Yosio Shimabukuro, "Analyzing the agricultural transition in Mato Grosso, Brazil, using satellite-derived indices", Applied Geography, 32(2):702-713, 2012.

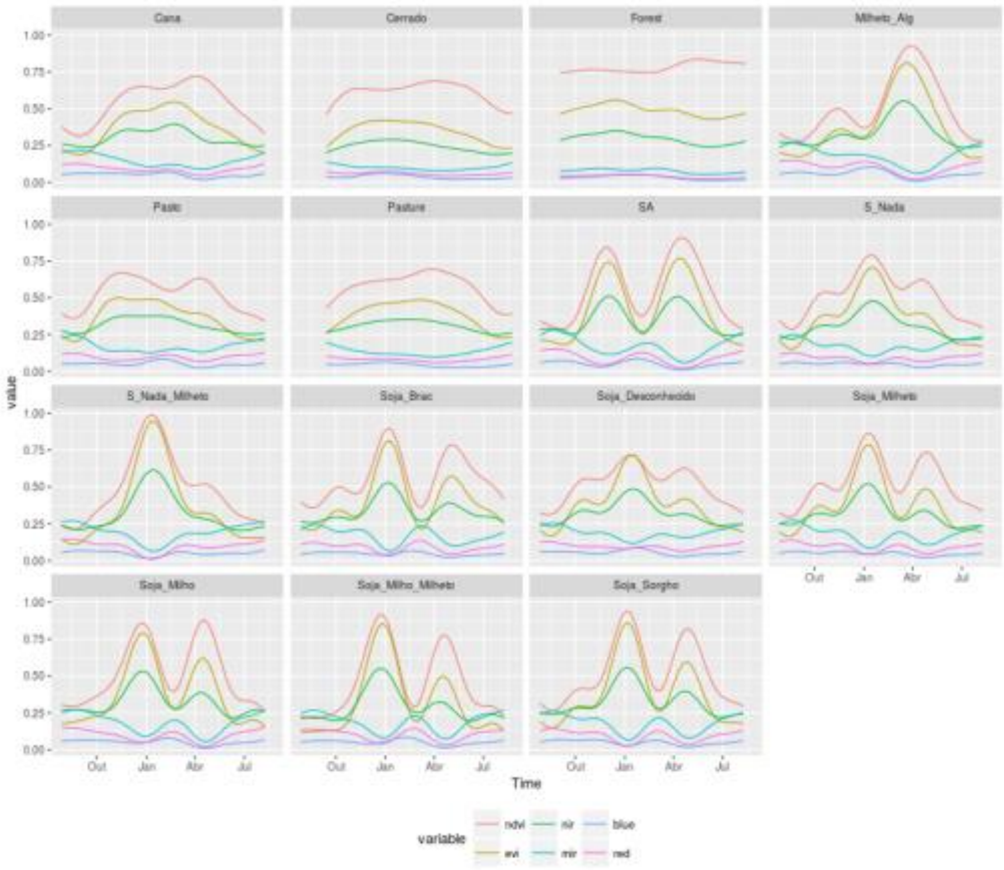

Figure 8 – Patterns of satellite image time series for the main crops produced in the Cerrado and Amazonia biomes of Brazil and for natural land covers (source: e-sensing project team).

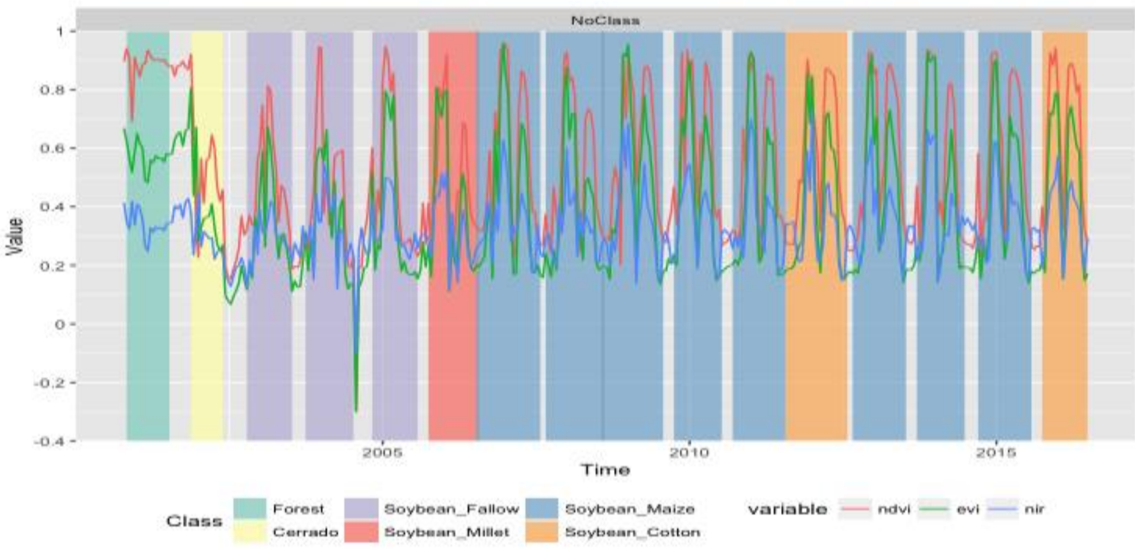

Figure 9 – Classification of land cover for location with longitude = -55.233 and latitude = -11.516 (source: e-sensing project team).

Futhermore, researchers from the project, combining expertise in data analysis (Prof. Dr. Leila Fonseca and Dr. Thales Körting) with expertise in agriculture (Prof. Dr. Ieda Sanches) carried out a study of using LANDSAT-8 image time series for crop mapping in the Cerrado [15]. They extracted parameters that describe the seasonality of signals using the TIMESAT software[6]. The variables obtained by TIMESAT were used for a data mining procedure using the Random Forest classifier. This study aimed to understand the potential and limitations of the TIMESAT software, which has been used by many authors in the literature to classify satellite image time series.

The authors found out that TIMESAT has been designed for find parameters for single yearly crops and that the software cannot deal well in situations of two or three crop seasons per year, which is quite common in Brazil. For example, Figure 10 shows the time series of the EVI vegetation index for an annual crop sample located into a triple cropping system of potato, bean and corn. The dotted line represents the time series outliers and null values removed and the thick line represents the double logistic filtered time series. Only one season was modelled by TIMESAT, corresponding to corn, planed in late September and harvested in early February.

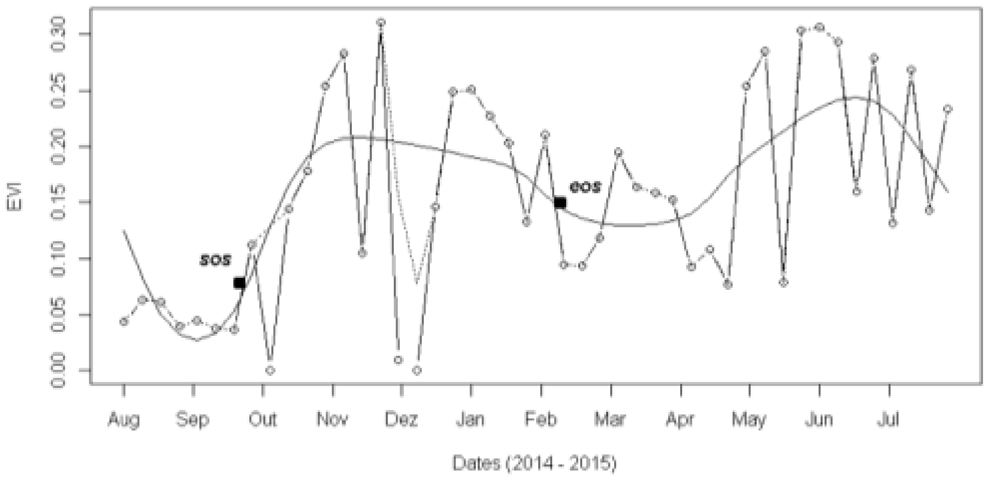

Figure 10 – LANDSAT-8 time series of an area with a a triple cropping system of potato, bean and corn (source: [15]).

[6] Jönsson, P., Eklundh, L. 2004. TIMESAT – a program for analyzing time-series of satellite sensor data. Computers & Geosciences. 30 (8), pp. 833-845.

The experiment performed in [14] was important to the project, so that we could better understand the potential and limitations of the TIMESAT software. Based on this experiment, we concluded that, at least for LANDSAT-class time series, the TWDTW method developed by the project team [3] has more capacity to detect double or triple cropping systems than TIMESAT. This hypotheses will be tested in Year 3 of the project, where the recently hired post-doc (Dr. Michelle Picoli) will carry out the mapping activities of Task 3.2 of the project.

4 Institutional support received in the period

The e-sensing project is hosted primarily by the Image Processing Division (DPI) of the National Institute of Space Research (INPE), with additional support provided by INPE's Remote Sensing Division (DSR). Both divisions report to the Earth Observation Directorate (OBT). During year 1 of the project, DPI gave substantial institutional support to the project, led by its head (Dr. Lubia Vinhas) as follows:

1. DPI/INPE hired, with additional funds from other projects, a full-time post-doc researcher (Dr. Eduardo Llapa) who is 100% dedicated to the project.
2. DPI/INPE also hired, with additional funds from other projects, a project support person (Ms. Denise Nascimento) who provides essential support for project management.
3. DPI/INPE also provides the IT infrastructure for hosting the data servers bought by the project with support from FAPESP, and for hosting the project's website (http://www.esensing.org).

The institutional support we are receiving from DPI/INPE is very good and fulfills the needs of the project.

5 Activities planned for project year 3 (January – December 2017)

5.1 Planned activities for WP 1 - Big Earth observation databases

5.1.1 Task 1.1 - Building and deployment of big Earth observation databases to support data analysis and use cases

Milestone M1.1.3 - Version 2 of the database for use cases in Brazil (month 36)

This milestone will consist of loading in the SciDB database the data for the use cases to be done in years 3 and 4. In this case, we expect to include a significant number of LANDSAT images to complement the MODIS data that has been already inserted in the SciDB array manager.

5.1.2 Task 1.2 – Extend SciDB for geographical data handling

Milestone M1.2.3 – Web service for SciDB server-side processing (month 36)

This milestone is focused on full implementation of *WTSCS* – web time series classification service, a service that allows the user to request the server-side processing of large Earth observation data sets. Using R as their primary tool for big data analytics, researchers will be able to scale up their methods, reuse previous work, and easily collaborate with their peers. Such a service would go beyond what is currently offered by OGC standards such as WPS and WCPS to allow progress on big EO data analytics.

5.2 Planned activities for WP 2 - Data analysis for big Earth observation data

5.2.1 Task 2.1 - Exploratory big data analysis

Milestone M2.1.3 – Interactive environment for collaborative analysis (month 36)

For this task, we intend to extend the exploratory data analysis environment for collaborative analysis to be open to other research groups in Brazil and abroad. The idea is to allow researchers to use the Web-based exploratory big data analysis tools (deliverable for Task 2.1 – see above) and the R-Big-EO-data analysis tools (deliverable for Task 2.2 – see above), so as to allow other groups to benefit from the project's results.

5.2.2 Task 2.2 – Space-time analysis of big Earth observation data for land change monitoring

Milestone M2.2.3: Version 1 of Big-EO space-time series R package (month 36)

This task aims to develop new methods for space-time analysis of big Earth observation data. In year 2, we have already produced new research results on satellite image time series analysis, which will be extended in Year 3. In Year 3, we will include tools to combine spatial and temporal information of Earth observation data in order to improve our analytical capability. This package will include tools for real-time detection of deforestation and with forest degradation.

5.3 Planned activities for WP 3 - Use case development

5.3.1 Task 3.1 - Specification and validation of tropical forest change alert methods and data

Milestone M3.1.3 Detection of clear cut and degradation (month 36)

In year 3, the recently hired Forestry post-doc (Rodrigo Begotti) will use the tools developed by the project to produce maps of forest clear-cut and forest degradation over large areas of the Amazon forest. We will then compare the results with data produced by the DETER and DEGRAD systems to assess the performance of these satellite image time series analysis methods.

5.3.2 Task 3.2 - Specification and Validation of Tropical Agriculture Monitoring Methods and Data

Milestone M3.2.3 Detection of area of soybeans, maize and sugarcane (month 24)

In Year 3, the recently hired Agricultural post-doc (Dr. Michele Picoli) we will use the *R-Big-EO time series analysis software* (developed in Task 2.2) to process large agricultural areas in Amazonia, Cerrado, and the state of São Paulo, with the specific tasks of mapping land cover associated to soybeans, maize and sugarcane. We will compare our results with ground truth data we have acquired and with results from IBGE (Brazil's Census Bureau).

6 Data Management Policy

We are following the policy we stated in the project proposal, as follows:

> *Our policy will be to deal with the databases and software created by this project as a resource to be shared with the Brazilian Earth Observation community. Thus, we will open the database after month 24 of the project to the community. We will encourage scientists to develop new data analysis methods and to use the methods and algorithms we will build to develop new applications. We will maintain the database accessible and updated for long-term use by the scientific community.*

Based on the results of Year 2, we decided to delay the opening of the project's database to the end of Year 3 (month 36). The reason is that we found out that the best way to allow outside researchers to use the database we have built is through a web-based collaborative working environment (see Task 2.1 above) combined with the *R-Big-EO time series analysis software* (see Task 2.2 above). A critical part of the infra-structure needed to open the archive is the *WTSCS* – web time series classification service, that allows users to request the server-side processing of large Earth observation data sets. This service will be developed during Year 3 of the project.

7 Final Remarks

The "e-Sensing" project has achieved important results in Year 2. We have improved out tools for satellite image time series analysis and developed a web-based exploratory data analysis environment. These results will allow the newly hired application post-docs in Forestry and Agriculture to produce new maps of land use change in Brazil. These maps will allow new insights into the trade-offs between environmental protection and agricultural production in Brazil.

As we remarked in the Year 1 Report, another important, although intangible result, was to have built an interdisciplinary approach to the problem of big Earth observation data handling. We held frequent seminars and workshops with the full project team, so that researchers could present their different viewpoints. It has been vitally important to have such discussions. As a result, all team members have deepened their understanding of the complex problem we will try to solve in the coming years.

8 Scientific papers published in 2016

(authors from research team are shown in red)

PAPERS PUBLISHED IN INTERNATIONAL JOURNALS

1. ISAQUE EBERHARDT, BRUNO SCHULTZ, RODRIGO RIZZI, IEDA SANCHES, ANTONIO FORMAGGIO, CLEMENT ATZBERGER MARCIO MELLO, MARKUS IMMITZER, KLEBER TRABAQUINI, WILLIAM FOSCHIERA, ALFREDO LUIZ, Cloud Cover Assessment for Operational Crop Monitoring Systems in Tropical Areas. *Remote Sensing*, vol.8(3), 2016. DOI:10.3390/rs8030219 (JCR: fator de impacto: 3.036).

2. MENG LU, EDZER PEBESMA, ALBER SANCHEZ, JAN VERBESSELT. Spatio-temporal change detection from multidimensional arrays: detecting deforestation from MODIS time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 117, pp. 227–236, 2016. DOI: 10.1016/j.isprsjprs.2016.03.007

3. VICTOR MAUS, GILBERTO CAMARA, RICARDO CARTAXO, ALBER SANCHEZ, FERNANDO RAMOS, GILBERTO QUEIROZ. A time-weighted dynamic time warping method for land use and land cover mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(8): 3729 – 3739, 2016. DOI: 10.1109/jstars.2016.2517118.

4. VIVIAN RENÓ, EVLYN NOVO, ISABEL ESCADA, Forest Fragmentation in the Lower Amazon Floodplain: Implications for Biodiversity and Ecosystem Service Provision to Riverine Populations. *Remote Sensing*, vol. 8, p.886, 2016.

5. TAÍSE PINHEIRO, ISABEL ESCADA, DALTON VALERIANO, PATRICK HOSTERT, FLORIAN GOLLNOW, Forest degradation associated with logging frontier expansion in the Amazon: the BR-163 region in Southwestern Pará, Brazil. *Earth Interactions*, 20(17):1-26, 2016.

PAPERS PUBLISHED IN BRAZILIAN JOURNALS

6. MARCIO AZEREDO, MIGUEL MONTEIRO, ISABEL ESCADA, KARINE FERREIRA, LUBIA VINHAS, TAÍSE PINHEIRO. Mineração de trajetórias de mudança de cobertura da terra em estudos de degradação florestal. *Revista Brasileira de Cartografia*, 68(4): 717-731, 2016.

7. ANDERSON SOARES, THALES KÖRTING, LEILA FONSECA. Improvements of the divide and segment method for parallel image segmentation. *Revista Brasileira de Cartografia*, v. 68(6), p. 1113-1122, n. 2016.

8. BRUNO SCHULTZ, ANTONIO FORMAGGIO, IEDA SANCHES, ISAQUE EBERHARDT, ALFREDO LUIZ, CLEMENT ATZBERGER. Classificação orientada a objetos em imagens multitemporais LANDSAT aplicada na identificação de cana-de-açúcar e soja. *Revista Brasileira de Cartografia*, v. 68(4), pp. 131-143, 2016.

9. LAERCIO NAMIKAWA, THALES KÖRTING, EMILIANO CASTEJON. Water body extraction from RapidEye images: An automated methodology based on Hue component of color transformation from RGB to HSV model. *Revista Brasileira de Cartografia*, v. 68(6):1097-1111, 2016.

10. THALES KÖRTING, KARINE FERREIRA, LUBIA VINHAS, MIGUEL MONTEIRO, GILBERTO QUEIROZ, Trends in GeoInformatics. *Revista Brasileira de Cartografia*, 68 (6): 1079-1086, 2016.

11. ALANA NEVES, HUGO BENDINI, THALES KÖRTING, LEILA FONSECA, Combining time series features and data mining to detect land cover patterns: A case study in northern Mato Grosso state, Brazil. RBC. *Revista Brasileira de Cartografia* v. 68(6): 1133-1142, n. 2016.

PEER-REVIEWED PAPERS IN SCIENTIFIC CONFERENCES

12. GILBERTO CAMARA, LUIZ FERNANDO ASSIS, GILBERTO QUEIROZ, KARINE FERREIRA, EDUARDO LLAPA, LUBIA VINHAS, "Big earth observation data analytics: matching algorithms with system architectures". In: Proceedings of the 5th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data - BigSpatial '16. Burlingname, USA, November 2016.

13. P. DIAZ, RAUL FEITOSA, IEDA SANCHES, GÍLSON COSTA, A Method To Estimate Temporal Interaction in A Conditional Random Field Based Approach For Crop Recognition. XXIII ISPRS Congress, Prague 2016. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, v. XLI-B7, p. 205-211. doi:10.5194/isprs-archives-XLI-B7-205-2016, 2016.

14. V. BRUM-BASTOS, BÁRBARA RIBEIRO, CAROLINA PINHO, THALES KÖRTING, LEILA FONSECA, Improvement evaluation on ceramic roof extraction using WorldView-2 imagery and geographic data mining approach. XXIII ISPRS Congress, 2016, Prague. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, v. XLI-B7, p. 883-889. doi:10.5194/isprs-archives-XLI-B7-883-2016.

15. HUGO BENDINI, IEDA SANCHES, THALES KÖRTING, LEILA FONSECA, ALFREDO LUIZ, ANTONIO FORMAGGIO, Using Landsat 8 image time series for crop mapping in a region of Cerrado, Brazil. In: XXIII ISPRS Congress, 2016, Prague. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, v.XLI-B8, p. 845 – 850. doi:10.5194/isprs-archives-XLI-B8-845-2016.

16. GILBERTO CAMARA, ADELINE MACIEL, VICTOR MAUS, LUBIA VINHAS, AND ALBER SANCHEZ. Using dynamic geospatial ontologies to support information extraction

from big earth observation data sets. In Ninth International Conference on Geographic Information Science (GIScience 2016), Montreal, CA, 2016. AAG.

17. THALES KÖRTING, LAERCIO NAMIKAWA, LEILA FONSECA, CARLOS FELGUEIRAS, How to effectively obtain metadata from remote sensing big data? GEOBIA 2016: Solutions and Synergies. Enschede, Netherlands. In Proceedings of GEOBIA 2016 (http://proceedings.utwente.nl/362/).

18. ANDERSON SOARES, THALES KÖRTING, LEILA FONSECA, First experiments using the Image Foresting Transform (IFT) algorithm for segmentation of remote sensing imagery. GEOBIA 2016: Solutions and Synergies. Enschede, Netherlands. In Proceedings of GEOBIA 2016 (http://proceedings.utwente.nl/362/).

19. CESARE GIROLAMO, A. PESSOA, THALES KÖRTING, LEILA FONSECA, Detecting Atlantic forest patches applying GEOBIA and Data Mining techniques. GEOBIA 2016: Solutions and Synergies. Enschede, Netherlands. In Proceedings of GEOBIA 2016 (http://proceedings.utwente.nl/362/).

20. LUIZ ASSIS, GILBERTO QUEIROZ, KARINE FERREIRA, LUBIA VINHAS, EDUARDO LLAPA, ALBER SANCHEZ, VICTOR MAUS, AND GILBERTO CAMARA. Big data streaming for remote sensing time series analytics using MapReduce. In XVIII GEOINFO (Brazilian Symposium on Geoinformatics), Campos do Jordão, SP, Brazil, 2016. Proceedings, pages 228-239 (www.geoinfo.info).

21. HUGO BENDINI, LEILA FONSECA, THALES KÖRTING, RENAN MARUJO, IEDA SANCHES, JEFFERSON ARCANJO, Assessment of a Multi-Sensor Approach for Noise Removal on Landsat-8 OLI Time Series Using CBERS-4 MUX Data to Improve Crop Classification Based on Phenological Features. In: XVIII GEOINFO (Brazilian Symposium on Geoinformatics), Campos do Jordão, SP, Brazil, 2016. Proceedings, pages 240 – 251 (www.geoinfo.info).

22. CESARE GIROLAMO, THALES KÖRTING, LEILA FONSECA, Assessment of texture features for Brazilian savanna classication: a case study in Brasilia National Park. In: XVIII GEOINFO (Brazilian Symposium on Geoinformatics), Campos do Jordão, SP, Brazil, 2016. Proceedings, pages 204-215, 2016 (www.geoinfo.info).

23. VITOR GOMES, LUCIANE SATO, GILBERTO QUEIROZ, LUBIA VINHAS, KARINE FERREIRA, Gerenciamento de nuvem de pontos em SGBD: avaliando a extensão PointCloud para PostgreSQL. In: XVIII GEOINFO (Brazilian Symposium on Geoinformatics), Campos do Jordão, SP, Brazil, 2016. Proceedings, pages 275-283, 2016 (www.geoinfo.info).

24. LORENA SANTOS, KARINE FERREIRA, GILBERTO QUEIROZ, LUBIA VINHAS. Spatiotemporal data representation in R. In: XVIII GEOINFO (Brazilian

Symposium on Geoinformatics), Campos do Jordão, SP, Brazil, 2016. Proceedings, pages 178-191, 2016 (www.geoinfo.info).

25. ALEXANDRO SILVA, LEILA FONSECA, THALES KÖRTING. Bayesian network model to predict areas for sugarcane expansion in Brazilian Cerrado. In: XVIII GEOINFO (Brazilian Symposium on Geoinformatics), Campos do Jordão, Brazil, 2016. Proceedings, pages 216-227, 2016 (www.geoinfo.info).

26. ROLF SIMÕES, GILBERTO QUEIROZ, KARINE FERREIRA, LUBIA VINHAS, GILBERTO CAMARA. PostGIS-T: towards a spatiotemporal PostgreSQL database extension. In: XVIII GEOINFO (Brazilian Symposium on Geoinformatics), Campos do Jordão, Brazil, 2016. Proceedings, pages 252-261, 2016 (www.geoinfo.info).

27. LUBIA VINHAS, GILBERTO QUEIROZ, KARINE FERREIRA, AND GILBERTO CAMARA. Web Services for Big Earth observation data. In: XVIII GEOINFO (Brazilian Symposium on Geoinformatics), Campos do Jordão, Brazil, 2016. Proceedings 166-177, 2016 (www.geoinfo.info).

PAPERS PRESENTED IN SCIENTIFIC CONFERENCES

28. GILBERTO CÂMARA, *"From VGI to CGI: Collaborative geographical initiatives as a basis for improved spatial information production"*. Paper presented at Vespucci Workshop 2016 on: *"Voluntary Geographic Information & Policy: managing, integrating and targeting crowdsourced information"*. Firenze, Italy, 2016.

DOCTORAL DISSERTATIONS

29. VIVIAN FRÓES RENÓ, Várzeas Amazônicas: Alterações da Paisagem e seus Impactos na Provisão de Serviços Ecossistêmicos e Bem-estar de Comunidades Riberinhas. Doctoral dissertation in Earth System Science, INPE 2016. ADVISOR: ISABEL ESCADA.

30. VAGNER LUÍS CAMILOTTI, Uso e Importância de Recursos Florestais Extrativistas em Comunidades Rurais na Amazônia e suas Relações com Aspectos Socioeconômicos Locais e Características da Paisagem. Doctoral dissertation in Earth System Science, INPE 2016. Advisor: ISABEL ESCADA.

31. VICTOR MAUS, Land use and land cover monitoring using remote sensing image time series. Doctoral dissertation in Earth System Science, INPE 2016. Advisor: GILBERTO CAMARA.

MASTER THESIS

32. ANIELLI ROSANE DE SOUZA, Economia e Natureza: Padrões de Uso e Cobertura da Terra Associados a Atividades Agropecuárias e Extrativistas de Comunidades do Sudoeste do Pará. MSc in Remote Sensing, INPE 2016. Advisor: ISABEL ESCADA.

ANNEXES

1. **Report of use of Technical Reserve and Complementary Benefits (in Portuguese)**

2. **Reports for Technical Training Scholarships:**

   2.1 Luiz Fernado Assis (2015/19540-0)

   2.2 Alber Sanchez Ipia (2016/16555-0)

3. **Reports for Postdoctoral Scholarships**

   3.1 Rodrigo Anzolin Begotti (2016/16968-2)

4. **Reports for Doctoral Scholarships**

   4.1 Rennan de Freitas Bezerra Marujo (2016/08719-2

5. **Initial Pages of Papers published**

6. **Copies of accecptance sheets for MSc and PhD thesis**

   6.1 Rennan de Freitas Bezerra Marujo (2016/08719-2

NOTE: The post-doc scholarship granted to Michelle Cristina Araújo Picoli (process 2016-23750-3) only started in 01.01.2017. Thus, there is no report of activities for year 2016.

ANEXO 1

RELÁTORIO SUCINTO DE UTILIZAÇÃO DE RECURSOS DA RESERVA TÉCNICA E BENEFÍCIOS COMPLEMENTARES

1. <u>DIÁRIAS E DESPESAS DE TRANSPORTE</u>

Foram utilizados recursos dessas rubricas para custear:

a) A participação do bolsista Hugo Bendini (*"Using landsat 8 image time series for crop mapping in a region of cerrado, Brazil")* e Victor Maus *("dtwSat: Time-Weighted Dynamic Time Warping for satellite image time series analysis in R"),* para participação no *XXIII ISPRS Congress* em Praga, República Tcheca.
b) A participação do pesquisador visitante, Clodoveu Davis Junior, em workshop no Brasil.
c) A participação da Dra. Karine Reis Ferreira Gomes na Conferência "*ACM Sigspatial 2016"* em *San Francisco, California, USA*: *"Big Earth Observation Data Analytics: Matching Requirements to System Architectures"*.
d) A participação do pesquisador principal do projeto, Dr. Gilberto Câmara Neto, no Workshop 2016 da Iniciativa Vespucci: *"Voluntary Geographic Information & Policy: managing, integrating and targeting crowdsourced information"* , em Florenca, Itália, com o trabalho: *"From VGI to CGI: Collaborative geographical initiatives as a basis for improved spatial information production"* e em reuniões de trabalho no Instituto de Geoinformática (IFGI) da Universidade de Münster, Alemanha, para discutir a preparação de um artigo cientifico vinculado ao projeto "e-sensing", a ser desenvolvido em conjunto com o prof. Edzer Pebesma (IFGI): *"Describing land trajectories using events"*.

2. <u>MATERIAL PERMANENTE</u>

Aquisição de monitores e quadro branco.

3. <u>MATERIAL DE CONSUMO</u>

Compra de baterias para no-breaks, cabos HDMI, teclados, mouses, cartuchos de toner, pen drive e fonte de carregador de notebook.

4. SERVIÇOS DE TERCEIROS

1. Contratação de instrutor para Curso de treinamento avançado de programação na linguagem R, sobre os pacotes de Análise Espacial e de Análise de Séries Temporais.
2. Pagamento de taxa de inscrição para participação em conferências e simpósios:
    - Dra. Karine Reis Ferreira Gomes: *"Big Earth Observation Data Analytics: Matching Requirements to System Architectures"*.
    - Dr. Gilberto Câmara Neto: *"Using dynamic geospatial ontologies to support information extraction from big Earth observation data sets".*
    - Rennan Marujo e Wanderson Costa (SIBGRAP).
    - Divulgações de bolsa de pós-doutorado do projeto em sites internacionais *(AAAS/ Science e ResearchGate).*
    - Compra de licença do Windows.
    - Conserto de notebook.
    - Seguro saúde para participação no *XXIII ISPRS Congress.*

ANNEXES

1. **Report of use of Technical Reserve and Complementary Benefits (in Portuguese)**

2. **Reports for Technical Training Scholarships:**

   2.1 Luiz Fernando Assis (2015/19540-0)

   2.2 Alber Sanchez Ipia (2016/16555-0)

3. **Reports for Postdoctoral Scholarships**

   3.1 Rodrigo Anzolin Begotti (2016/16968-2)

4. **Reports for Doctoral Scholarships**

   4.1 Rennan de Freitas Bezerra Marujo (2016/08719-2

5. **Initial Pages of Papers published**

6. **Copies of accecptance sheets for MSc and PhD thesis**

NOTE: The post-doc scholarship granted to Michelle Cristina Araújo Picoli (process 2016-23750-3) only started in 01.01.2017. Thus, there is no report of activities for year 2016.

ANEXO 1

RELÁTORIO SUCINTO DE UTILIZAÇÃO DE RECURSOS DA RESERVA TÉCNICA E BENEFÍCIOS COMPLEMENTARES

1. DIÁRIAS E DESPESAS DE TRANSPORTE

Foram utilizados recursos dessas rubricas para custear:

a) A participação do bolsista Hugo Bendini (*"Using landsat 8 image time series for crop mapping in a region of cerrado, Brazil")* e Victor Maus *("dtwSat: Time-Weighted Dynamic Time Warping for satellite image time series analysis in R")*, para participação no *XXIII ISPRS Congress* em Praga, República Tcheca.
b) A participação do pesquisador visitante, Clodoveu Davis Junior, em workshop no Brasil.
c) A participação da Dra. Karine Reis Ferreira Gomes na Conferência "*ACM Sigspatial 2016"* em *San Francisco, California, USA: "Big Earth Observation Data Analytics: Matching Requirements to System Architectures"*.
d) A participação do pesquisador principal do projeto, Dr. Gilberto Câmara Neto, no Workshop 2016 da Iniciativa Vespucci: *"Voluntary Geographic Information & Policy: managing, integrating and targeting crowdsourced information"* , em Florença, Itália, com o trabalho: *"From VGI to CGI: Collaborative geographical initiatives as a basis for improved spatial information production"* e em reuniões de trabalho no Instituto de Geoinformática (IFGI) da Universidade de Münster, Alemanha, para discutir a preparação de um artigo cientifico vinculado ao projeto "e-sensing", a ser desenvolvido em conjunto com o prof. Edzer Pebesma (IFGI): *"Describing land trajectories using events"*.

2. MATERIAL PERMANENTE

Aquisição de monitores e quadro branco.

3. MATERIAL DE CONSUMO

Compra de baterias para no-breaks, cabos HDMI, teclados, mouses, cartuchos de toner, pen drive e fonte de carregador de notebook.

4. SERVIÇOS DE TERCEIROS

1. Contratação de instrutor para Curso de treinamento avançado de programação na linguagem R, sobre os pacotes de Análise Espacial e de Análise de Séries Temporais.
2. Pagamento de taxa de inscrição para participação em conferências e simpósios:
    - Dra. Karine Reis Ferreira Gomes: *"Big Earth Observation Data Analytics: Matching Requirements to System Architectures"*.
    - Dr. Gilberto Câmara Neto: *"Using dynamic geospatial ontologies to support information extraction from big Earth observation data sets".*
    - Rennan Marujo e Wanderson Costa (SIBGRAP).
    - Divulgações de bolsa de pós-doutorado do projeto em sites internacionais *(AAAS/ Science e ResearchGate).*
    - Compra de licença do Windows.
    - Conserto de notebook.
    - Seguro saúde para participação no *XXIII ISPRS Congress.*

ANNEX 2.1

Integration between R, TerraLib and SciDB

Luiz Fernando Ferreira Gomes de Assis

Coordinator: Gilberto Câmara

Co-Advisor: Karine R. Ferreira

Project Number: 2015/19540-0

INPE - São José dos Campos
December, 2016

1 Introduction

This scientific report aims to describe the whole progress of the project entitled *Integration between SciDB, TerraLib and R*[3] developed by the awarded technical training fellow Luiz Fernando Ferreira Gomes de Assis[4]. This project started on the 1st of December 2015 and ended on the 31th of November 2016 and is a part of the e-Science Program [5], a thematic grant funded by São Paulo Research Foundation. The research addresses how the scientific community can use e-Science methods and techniques to improve the extraction and analysis of land use and land cover change information from big Earth Observation data sets in an open and reproducible way. This project is called *e-Sensing: Big Earth Observation Data Analytics for Land Use and Land Cover Change Information* and is coordinated by Prof. Dr. Gilberto Câmara Neto[6]. The e-Sensing team members consists of MSc. and PhD students, as well as Postdocs, and Researchers. Its excellence and its published papers are related to Geoinformatics and GIScience.

The main goal to develop this project is in the enhancement of the performance to integrate an statistical environment with a multidimensional array database in order to provide imaging routines for land use classification. The most used open source statistical environment is **R** since it is easily extensible through a substantial set of statistical functions and packages. Although the **R** environment provides a wide variety of graphical and statistical tools, it still has main memory and performance limitations when executing routines with large volumes of remote sensing images compared to other languages.

In this sense, the integration between **R** and multidimensional array databases such as SciDB can offer processing and analysis in the data server environment minimizing the data transfer between client and server by means of an interface with the **R** environment. For all the aforementioned reasons, we developed an interface between the **R** data analysis language, the SciDB multidimensional array database, and a library able to handle data stored in a PostGIS database such as TerraLib library. This architecture based on open-source tools was

[3]http://www.bv.fapesp.br/pt/bolsas/161826/integracao-entre-scidb-terralib-e-r/

[4]http://www.bv.fapesp.br/pt/pesquisador/669236/luiz-fernando-ferreira-gomes-de-assis/

[5]http://www.fapesp.br/8436

[6]http://www.bv.fapesp.br/pt/pesquisador/997/gilberto-camara-neto/

evaluated to assess how it meets the needs of Earth Observation (EO) scientists taking into consideration related works and a set of criteria to build researcher-friendly architectures for big EO data analysis.

The remainder of this scientific report is organized as follows. Section 2 contains two main studies performed to accomplish the software development of this project. Section 3 describes deeply how the software was developed. Section 4 comprises important applications case studies for evaluating the software development. Section 5 contains a description of the complementary activities to develop this project. Section 6 consists of a conclusion about a description and an evaluation of the institutional support. Finally, the papers written for better documenting this project are included in the appendices.

2 A driven-study for the software development

Before starting each project step of software implementation, we performed driven-studies about important topics of the project. Some of them are described in more detail in this section.

2.1 TWDTW: A Time-Weighted Dynamic Time Warping method for land use and land cover mapping

In this scientific report, we aim to focus on a supervised method for land use cover classifications. This because, it is still difficult to analyze the great annual variability of crop pheonological cycles since they depend on climate conditions and the interest of land managers. We also need to concern about irregularly sampling and noisy data, that is, the quality of the training pixels used to define the classes and the quality of the image used as input for classification. One method was recently proposed with the aiming of calculating an optimal alignment between mid-term and long-term time series, by warping the time axis iteratively through a temporal constraint, that can avoid either the inconsistent matching or alignment dependent on the seasons (see Figure 1) [2].

This algorithm is called Time Weighted Dynamic Time Warping (TWDTW) and is an adaptation of the Dynamic Time Warping (DTW) algorithm [3]. It aims to add a time-weight in the similarity measure provided by DTW algorithm, by providing flexible time segments when dealing with phenomenon with strong time dependency that influence the comparison of mid-terms such as the temporal patterns and long-term time series such as those provided by satellite imagery.

For calculating the similarity measure, we need firstly to create a distance matrix, considering it as a terrain, where white regions are the valleys and dark regions are the peaks. What we want to know is the optimum path leaving from the lower left corner position to get to the upper right corner position by avoiding the peaks and taking as much valleys as it is possible. Then, we build artificial boundaries to facilitate the first few decisions and allow the path to be built holding positions choices inside the matrix. After that, for each position, we choose north, north east and east as a next step, taking into consideration the less expansive position to get close to the upper right corner (see Figure 2).

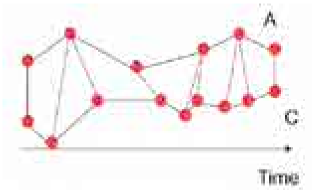

Figure 1: Two time-series with different lengths.

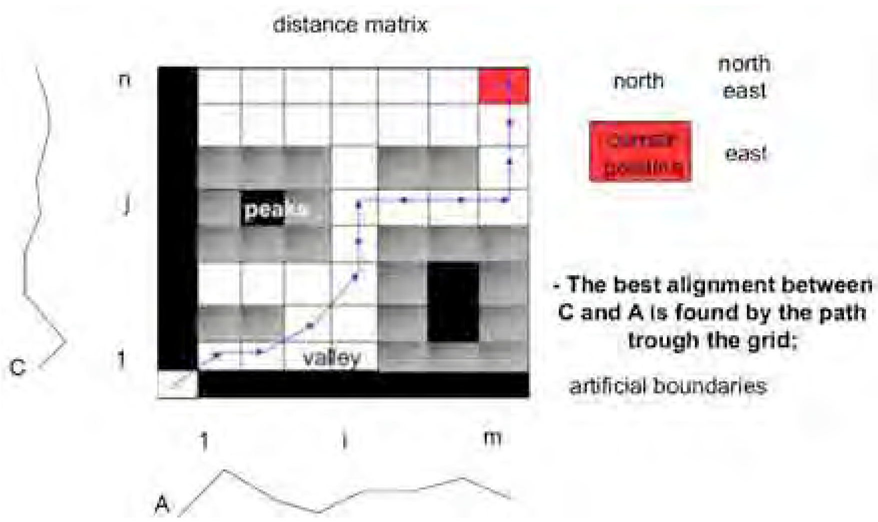

Figure 2: Dynamic Time Warping.

2.2 Manipulating Time Series in **R**

Since the objects contained in the **R** base package are not built to work efficiently with time series, a set of packages, objects and classes have been created for manipulating them. Figure 3 gives an overview of how time series can be used in **R**.

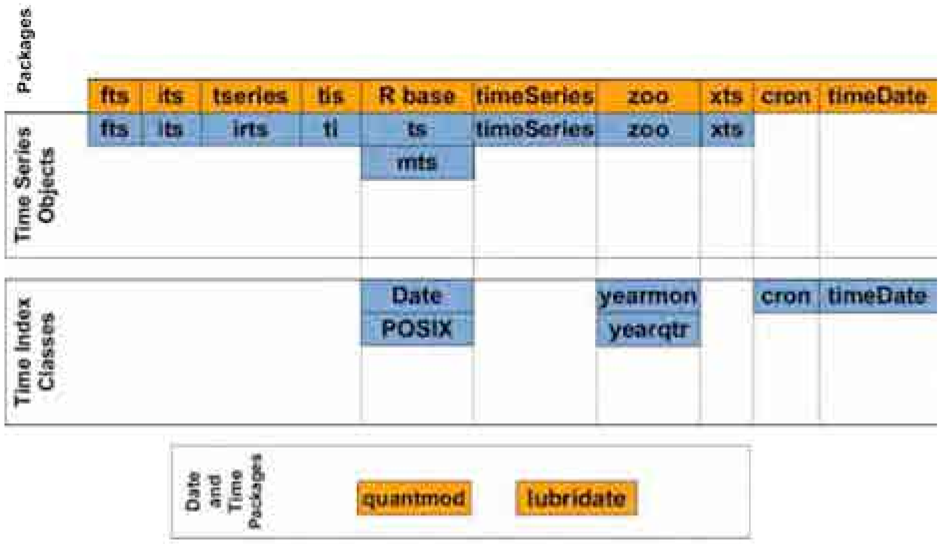

Figure 3: R packages, objects and classes for time series.

The following packages allow users to manipulate time series: *fts* package provides fast operations for time series; the *its*, *irts* and *tseries* packages are better fitted for irregularly spaced time series; the *tis* package can support the management of time index objects; the *stats* package are suitable for representing univariate and multivariate regularly spaced time series. However, the problem about these representations is that they cannot adequately represent more general irregularly spaced non-calendar time series.

For this, most of the users have used *timeSeries*, *zoo* and *xts* packages. The time series data represented by *timeSeries*, *zoo* and *xts* objects contain the time index and the data, both stored in a vector of date-time and a rectangular data object respectively. *zoo* objects are very flexible and were designed to deal with arbitrary ordered time index, while *xts* objects are represented by formal time-based classes for indexing, internal properties, and user-added attributes.

The **R** base package also helps to deal with date to represent data on discrete dates without considering the time of the day, while *POSIXct* and *POSIXlt*, derived from the POSIX system, allow users to represent dates and times controlling time zones, which is specially important for intra-day transactions-level. *chron* classes are similar to the *POSIXt* classes but not used so often as them. *quantmod* package is a prototyping environment designed to assist in the development, testing and deployment of statistically based trading models. Lastly, *lubridate* package provides a smart parsing to work with dates and times in **R**. The **R** package dependencies presented here can be seen in Figure 4.

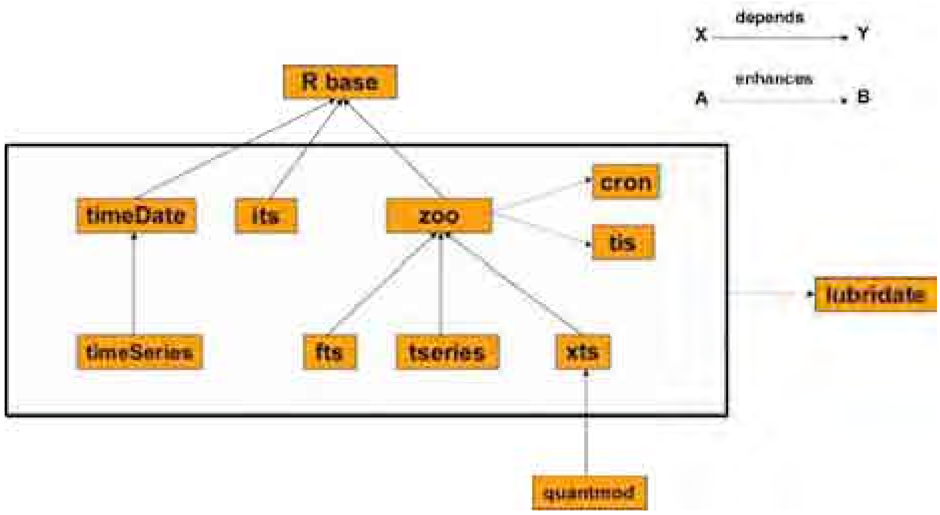

Figure 4: **R** packages dependencies in time series.

2.3 Multidimensional Array Database (SciDB)

Multidimensional array databases store data over a cluster composed of nodes that are capable of performing multiple aggregate complex instructions independently (see Figure 5). The coordinator node is responsible for the query execution and optimising, while the instance nodes are responsible for the query processing. Specifically, SciDB has been used due to their design to be extensible and parallel. SciDB's parallelism is performed by means of data partitioning into chunks optimized by an overlapping strategy [4]).

[7] http://www.dpi.inpe.br/˜gribeiro/lib/exe/fetch.php?media=wiki:cursos:big_geospatial_data_com_scidb_-_scidb.pdf

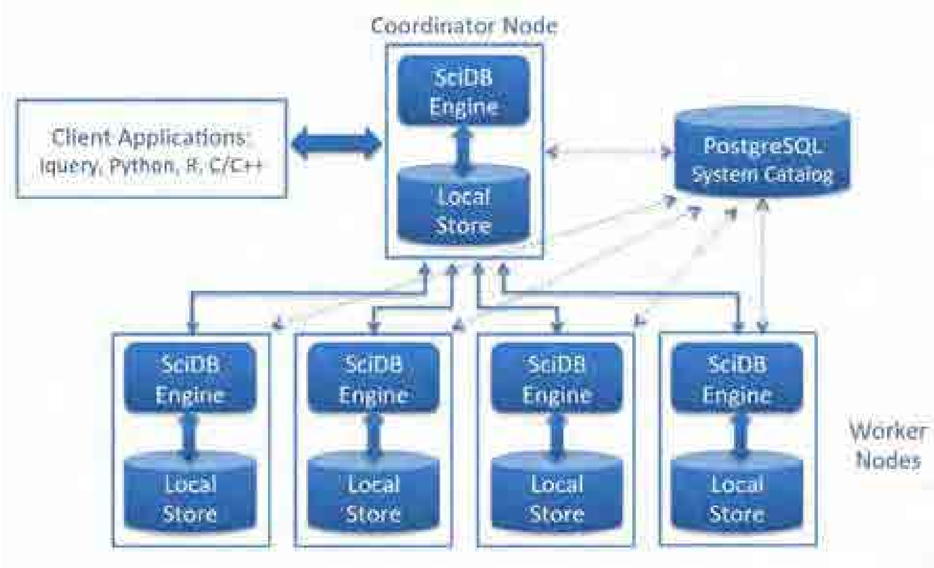

Figure 5: SciDB architecture[7].

SciDB has been used as a solution-based for loading and analyzing NASA's data [5]. However, a set of additional tools is necessary to the loading preprocess responsible for converting different data formats to the SciDB representation. SciDB also needs to be extended for spatio-temporal interfaces and field data type [6], as well as more complex statistical analysis. Although analytic algorithms have been built to enrich multidimensional array databases, they still lag far behind statistical software packages such as those presented in the CRAN repository.

SciDB exploits the matrix data model to provide an efficient chunks-based storage mechanism and vertical partitioning by array attributes. In addition, it has two high-level query languages: one named Array Query Language (AQL) similar to SQL, and another called Array Functional Language (AFL) related to the form of function composition of functional languages.

2.4 BFAST: Breaks For Additive Season and Trend

BFAST is an algorithm that stands for Breaks for Additive Season and Trend aiming to integrate, detect and characterize the decomposition of and structural changes within time series components (trend, seasonal, and remainder) [7]. It has also been used to analyze satellite imagery time series in near real-time based on fitted piecewise linear models, handling missing values without interpolation [8]. An exemplary deployment of BFAST method in **R** is depicted in Figure 6.

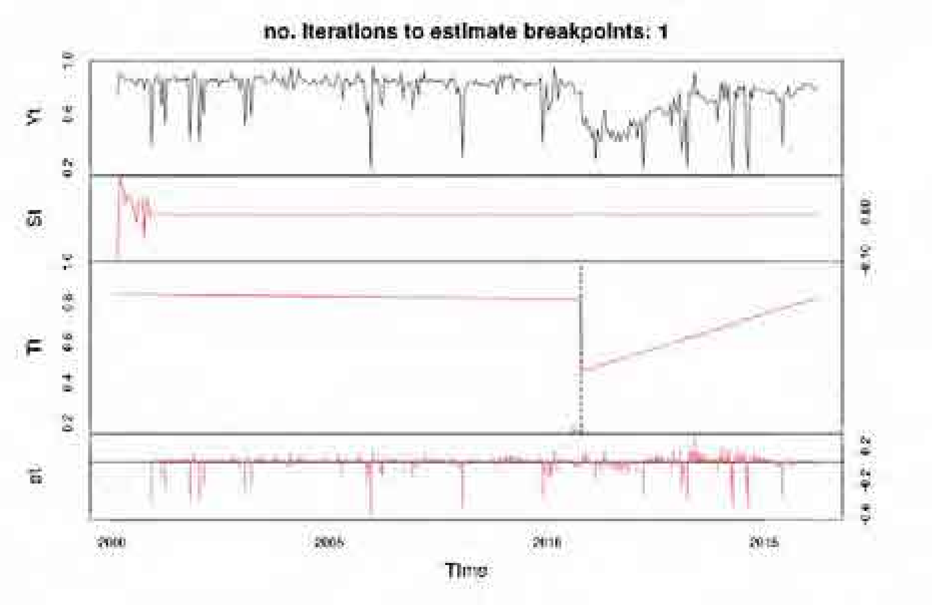

Figure 6: Breaks for Additive Season and Trend.

The BFAST package has three main methods: BFAST01, BFAST and BFAST monitor. BFAST01 aims at checking only one major break in the time series. BFAST aims at detecting breaks in the seasonal and trend component of a univariate time series, while BFAST monitor aims at detecting changes based on a model for stable historical behaviour abnormal within newly acquired data.

3 Software Development

In this section, we discuss the main software implementation and development performed by the technical training fellow. We mainly focus on the step-by-step **R**-SciDB integration and its enhancements developed during the project.

3.1 R_exec (First contacts with R-SciDB integration)

Although database technologies have become popular, they are still not able to completely manage and analyze scientific data due to their data model inefficiency and storage strategies problems to meet the fundamental requirements of the scientific community for big data. To overcome this problem, take the best of combining two worlds that include programming flexibility and processing efficiency can help to reduce the data movement and the communication overhead [9]. These approaches aim to scale deep analytic methods for massive datasets by exploiting the parallelism of database technologies in an analyst-friendly environment [10, 11].

In this sense, an alternative way of integration that invokes deep analytics methods scripts inside databases queries were proposed [1, 12]. The **R**-SciDB integration facilitates data exchange between both worlds, since either users may avoid using only built-in SciDB operations, or moving big data by means of stdin and stdout of **R**. For this, an older SciDB library plugin called r_exec was firstly considered as an alternative to invoke **R** scripts independently over all chunks in each SciDB instance using queries for arrays as input (see Figure 7).

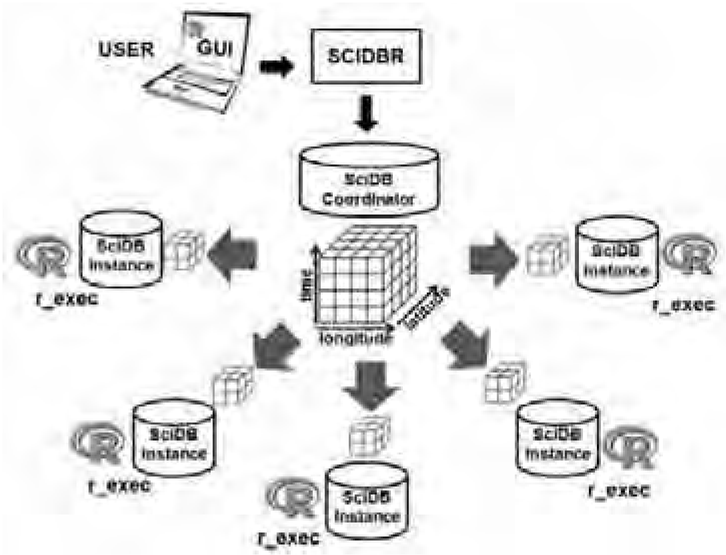

Figure 7: Older **R**-SciDB integration [1].

The source code available required minor modifications to work with SciDB 14.12, the version available in our cluster, mainly changes in the r_exec physical and logical settings code were necessary. We added an *using namespace std* at the initial of the code, and an *std* every time a

[7] https://github.com/Paradigm4/r_exec

method from it was called. This because *boost* and *std* libraries were confused for some of the methods. We also exchange the *MAX_COORDINATE* constant to the *CoordinateBounds::getMax()* one, compiling all the Makefile after all of this using CMAKE version 11.

3.2 TerraLib GeoWeb Services Server-Side

TerraLib GeoWeb Services (TWS) is an open source platform for the development of geospatial web services that handles, share and process Earth Observation data on the web. It is a simple and more effective API, when comparing with standard services, for accessing and analyzing big EO data [13]. This viewpoint is highlighted by the implementation of a Web Time Series Service (WTSS) that offers remote sensing time series. WTSS contains coverages represented by three dimensional arrays with spatial and temporal dimensions. It also provides three operations: list_coverages, describe_coverage and time_series.

The list_coverages operation returns a list of all available coverages in the service, describe_coverage operation returns the metadata of a given coverage, and time_series operation consists of querying the database for a list of values for a given location and time interval. The results of the operations are provided as JSON documents, while the service is implemented in C++.

Within the WTSS implementation, there is a parsing of the initial and final time in which users are interested. A timeline range component is responsible for checking whether the start and end time match to the period of the specified coverage. Since users can also define the searched attributes, it is also required to manage the dataset values. We also need to convert the latitude and longitude provided by the user to the one specified in the database using the array projection system of TerraLib.

3.3 **R** Client API for Web Time Series Service

For handling remote sensing imagery as time series, we developed an **R** client-side package of a lightweight Web Time-Series Service (WTSS). wtss.R can be used as an input for a high variety of methods that aims to analyze satellite imagery time series. The **R** client is very easy to use,

[7] https://github.com/e-sensing/tws/tree/master/src/tws/wtss

we only need firstly to create the connection to the server-side. After that, coverages can be listed as an exploration of arrays in SciDB. Since the time series are provided in *zoo* objects (indexed ordered observations) by wtss.R, a coercion is still required when dealing with other methods that requires other objects types. A simple example of how installing, loading and using wtss.R package in **R** is described in Listing 1.

Listing 1: A simple example of using wtss.R

```
devtools::install_github("luizassis/wtss.R")
library(wtss.R)
# create a WTSS connection
ts_server = WTSS("http://www.dpi.inpe.br/tws/wtss")

# get the list of coverages provided by the service
coverages = listCoverages(ts_server)

# get the description of the second coverage
cv = describeCoverage(ts_server,coverages[2])

# get a time series
ts = timeSeries(ts_server,
                names(cv),
                cv[[1]]$attributes$name,
                latitude=-10.408,
                longitude=-53.495,
                start="2000-02-18",
                end="2016-01-01")

plot(ts[[1]]\$attributes[,1],
     main="Pixel Center Coordinates Time-Series",
     xlab="Time",
     ylab="Normalized Difference Vegetation Index")
```

3.4 SciDB Streaming vs Hadoop Streaming

Besides r_exec plugin, SciDB started offering an operator that streams data sets for **R** scripts. In order to use this operator called streaming, we firstly need to share the **R** scripts so that every machine can "see" them to execute. The streaming operator considers SciDB arrays as input and output. It also has as parameters a comma-separated list of column SciDB types and a comma-separated output column names with the same lengths, depending on the specified type. Figure 9 represents SciDB streaming architecture, which is similar to the r_exec.

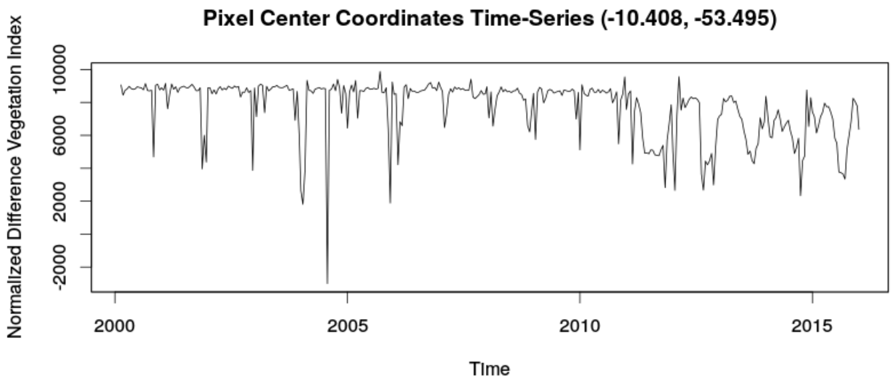

Figure 8: Vegetation index (ts time series).

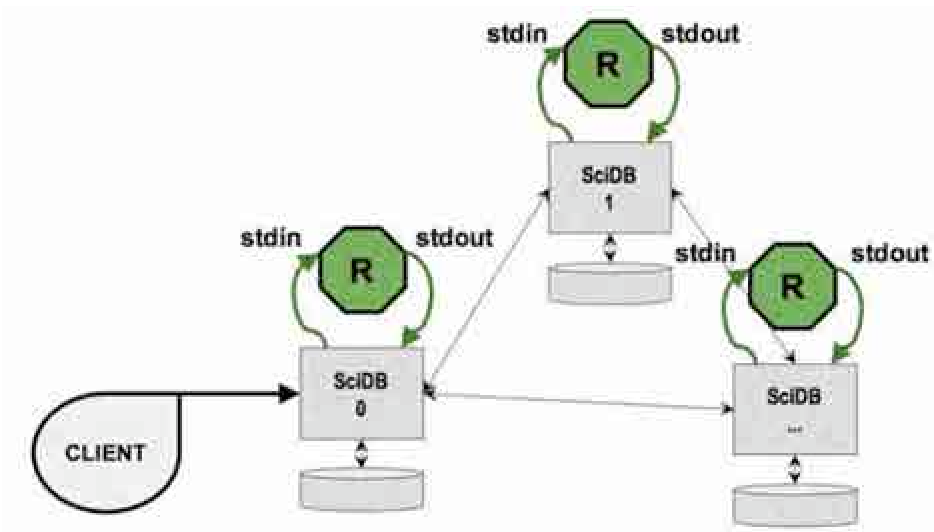

Figure 9: SciDB streaming architecture[8].

[8]https://github.com/Paradigm4/stream

SciDB streaming is also very similar to the Hadoop streaming, but it does not implement MapReduce model. Unlike SciDB streaming, Hadoop streaming needs an adaptation of the input and output datasets to sets of key/value pairs, and a manually setting to splitting the data into multiple nodes. With this, we loose the generality in terms of researcher-friendliness, since multidimensional array databases are easier to work with than map/reduce environments. More details about our experiments comparing SciDB Streaming and Hadoop Streaming can be seen in Appendix A and B.

4 Application Case Study

During this project we developed and enhanced the **R**-SciDB integration by means of the components described in the previous section. We tested them using two different application case studies: deforestation detection and land use classification.

4.1 Deforestation Detection: BFAST monitor extension

In our first application case study, we performed a set of experiments using the **R** BFAST package, which aims at detecting deforestation and phenological change, and monitoring forest health in large areas. Our initial tests consisted of checking how the overhead of these tools affected this kind of processing although we initially know this integration has a stable, adequate and linear performance even when the amount of information increase with the time. Our results showed that the limitation of the performance is upon to the hardware capabilities infrastructure in terms of storage and computation power.

The idea here to extend BFAST monitor is very simple. As previously mentioned, this algorithm requires the data, the start of the stable history period and the start of the monitoring period. The start of the stable history period is always the same, that is, we choose here the first observation of the analyzed long-term time series. On the other hand, the start monitoring period changes with each new acquired data and is constantly decreased until we find a break. It is important to mention that BFAST monitor only accepts time series with at least nine observations.

In an exemplary situation with a time series with ten observations, the start stable history period is the first observation, while the start monitoring period is the tenth observation at first, but decreases each new iteration. Since BFAST monitor only accepts more than nine observation, the algorithm stops in the next iteration. After that, when a new observation is acquired, the start monitoring period starts at the eleventh observations and decreases until the tenth observation if no break is found. The Figure 10 depicts an overview of the BFAST monitor extension, while the Listing 2 contains the algorithm.

Listing 2: BFAST monitor extension algorithm

```
time_series <- ts(data, freq, start)
i = length(time_series)

while(i > 9)
{

  bf = bfastmonitor(time_series,
                    start=time(time_series)[i],
                    history=time(time_series)[1])

  if(!is.na(bf$breakpoint))
  {
    points_break <- bf$breakpoint
    break
  }
  i = i - 1
}
```

4.2 Land Use Classification: TWDTW Evaluation

To organize the broad research topic of land use and land cover classification, we develop here a remote sensing time series approach using a statistical environmental and a multidimensional array database. Based on an ad hoc literature review [14], we design our approach based on a workflow with four main features for land use analytics: the initial definitions, the sampling, the classification and the analysis (Figure 11a).

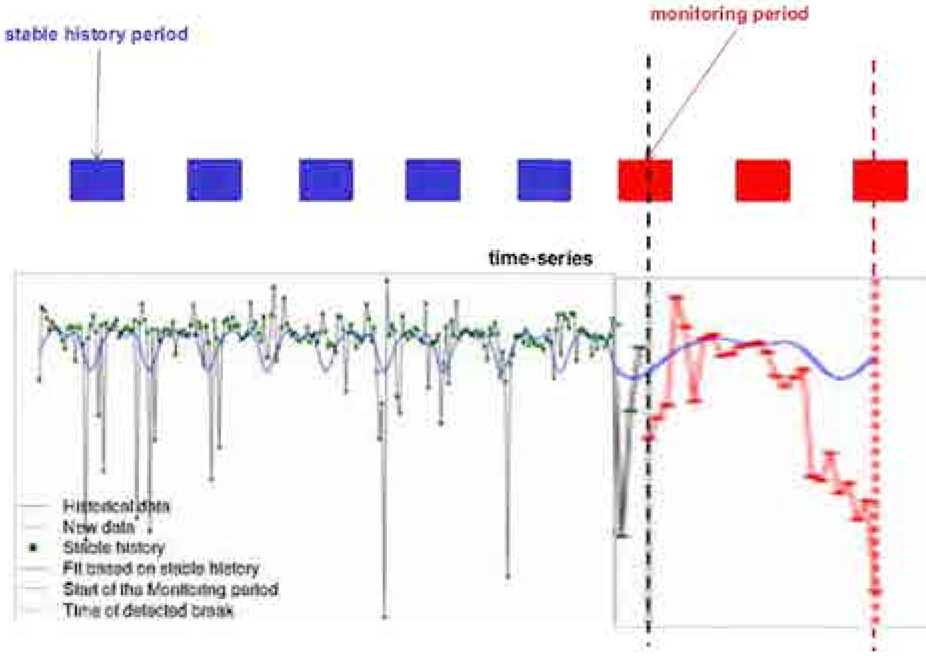

Figure 10: BFAST monitor extension overview.

Additionally, several relevant sub-features were included within each feature. The initial definitions include the spatial assessment unit, the sources of reference data, the reference labeling protocol and the agreement definitions. The sampling consists of the stage for selecting a subset of the smallest spatial units that help the classification methods. It includes the interpretation and quality control, phenology characterization and the homogeneity measurement. For a while, we are just considering supervised methods for classification, specifically, the classification results provided by the TWDTW method **R** package, also known as dtwSat.

Our analysis aims at measuring by means of estimating accuracy and class areas using data mining techniques. The workflow design and the feature layers were designed to reflect the land cover classification evaluation in order to conceptualize our analysis in a new way and to effectively and efficiently assist scientists in quickly grasp-ing the topics at a glance by only taking into considerations features that were discussed in this project.

References

[1] M. Lu, E. Pebesma, A. Sanchez, and J. Verbesselt, "Spatio-temporal change detection from multidimensional arrays Detecting deforestation from MODIS time series," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. March, 2016.

[2] V. Maus, G. Câmara, R. Cartaxo, A. Sanchez, F. M. Ramos, and G. R. de Queiroz, "A time-weighted dynamic time warping method for land-use and land-cover mapping," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (J-STARS)*, 2015.

[3] V. Maus, G. Câmara, R. Cartaxo, F. M. Ramos, A. Sanchez, and G. Q. Ribeiro, "Open boundary dynamic time warping for satellite image time series classification," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 3349–3352, IEEE, 2015.

[4] M. Stonebraker, P. Brown, D. Zhang, and J. Becla, "Scidb: A database management system for applications with complex analytics," *Computing in Science & Engineering*, vol. 15, no. 3, pp. 54–62, 2013.

[5] G. Planthaber, M. Stonebraker, and J. Frew, "Earthdb: scalable analysis of modis data using scidb," in *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*, pp. 11–19, ACM, 2012.

[6] G. Camara, M. J. Egenhofer, K. Ferreira, P. Andrade, G. Queiroz, A. Sanchez, J. Jones, and L. Vinhas, "Fields as a generic data type for big spatial data," in *Geographic Information Science*, pp. 159–172, Springer, 2014.

[7] J. Verbesselt, R. Hyndman, G. Newnham, and D. Culvenor, "Detecting trend and seasonal changes in satellite image time series," *Remote sensing of Environment*, vol. 114, no. 1, pp. 106–115, 2010.

[8] J. Verbesselt, A. Zeileis, R. Hyndman, and M. J. Verbesselt, "Package 'bfast'."

[9] P. Leyshock, D. Maier, and K. Tufte, “Agrios: A hybrid approach to big array analytics,” in *Big Data, 2013 IEEE International Conference on*, pp. 85–93, IEEE, 2013.

[10] R. Integrating, “Bridging two worlds with rice,” in *Proceedings of the VLDB Endowment*, 2011.

[11] S. Das, Y. Sismanis, K. S. Beyer, R. Gemulla, P. J. Haas, and J. McPherson, “Ricardo: Integrating R and Hadoop,” in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pp. 987–998, ACM, 2010.

[12] G. R. d. Queiroz, K. R. Ferreira, L. Vinhas, G. Camara, R. W. d. Costa, R. C. M. d. Souza, V. W. Maus, and A. Sanchez, “WTSS: um serviço web para extração de séries temporais de imagens de sensoriamento remoto,” in *Proceeding of the XVII Remote Sensing Brazilian Symposium*, pp. 7553–7560, 2015.

[13] L. Vinhas, G. Ribeiro, K. R. Ferreira, and G. Câmara, “Web services for big earth observation data,” in *Proceedings of the 17th Brazilian Symposium on GeoInformatics*, 2016.

[14] P. Olofsson, G. M. Foody, M. Herold, S. V. Stehman, C. E. Woodcock, and M. A. Wulder, “Good practices for estimating area and assessing accuracy of land change,” *Remote Sensing of Environment*, vol. 148, pp. 42–57, 2014.

National Institute for Space Research

Image processing division

ANNEX 2.2

Tools for satellite image management in array databases

Yearly Scientific Report of the Regular-Support Research Project,
funded by the São Paulo Research Foundation.

Projeto FAPESP #2016/03397-7

Responsible Researcher: Alber Hamersson Sánchez Ipia

São José dos Campos, 22 de Fevereiro de 2017

1 Summary of the project

This project is part of the the *e-sensing* FAPESP project (grant 2014/08398-6). Specifically, it corresponds to the task *Building and deployment of big Earth observation databases to support data analysis and use cases* [1].

The e-sensing project addresses the scientific question *How can we use e-science methods and techniques to substantially improve the extraction of land use and land cover change information from big Earth Observation data sets in an open and reproducible way?* Currently, scientists do not take advantage of the full potential of the freely available satellite images. Instead, they produce land cover maps taking either a single or at most two time references. As a result, the big data sets produced by remote sensing satellite are underemployed. The e-sensing project is about conceiving, building, and deploying a new type of knowledge platform for organizing, accessing, processing and analyzing big Earth observation data [1].

The e-sensing project is framed by the fast land cover and land use change and its global consequences on the Earth's systems. As the human population grows, it also grows the demand of resources from the environment. These demands are satisfied by changing the use — and in consequence, the cover — of the Earth. As the surface of the Earth is finite, mankind development is jeopardized [2, 3].

Human beings have been changing Earth surface to satisfy their needs for millennia. However, it is just until recently when we acquire the capacity to collect massive amounts of data to study the changes of Earth's surface. Satellite images date back to the 70s; they contain detailed traces of human development for the last 50 years and they are publicly available [4].

However, scientist are unable to use all the available images in their analyses because processing such volume of data demands large hardware resources, new software tools, and sound analysis techniques. These issues and requirements are known as the *data deluge* or more commonly as *big data* [5, 6, 7].

The current satellite image distribution model is based on files. These files have their own formats and access interfaces. This distribution model had led to problems such as data duplication and the inability to track the files used for specific analysis. This contributes to the already existent reproducibility crisis in science; specifically, the data used for Earth Observation analysis are either unavailable or just too large for independent result validation [8, 9].

Independently of format or interface, images are stored and manipulated using an array pattern. Arrays are well-known structures for scientists, take for example the Network Common Data Form (NetCDF) and Hierarchical Data Format (HDF). Recently, computer scientists have mixed the array abstraction with the features of relational databases into what has been called *array databases*. Array databases add versioning, scalability, and fail-tolerance capabilities to the array abstraction [10, 11, 12, 13, 14].

As a solution to the aforementioned issues, the e-sensing project proposes an open-source knowledge platform for big Earth Observation data. Such platform would provide an homogeneous interface to organize, access, process, and analyze spatiotemporal data using by means of array databases. In this way, scientists will analyse and test their hypotheses using data with a larger extents and finer resolutions than before. Likewise, this platform will enable reproducibility as any scientist can reproduce anyone else results using the same data and interfaces [1].

The e-sensing project team chose SciDB as the array database to support the proposed knowledge platform. SciDB is an open source array database optimized for big data and analytics. SciDB is developed and maintained by the Massachusetts Institute of Technology. SciDB splits and distributes data among several servers following a *shared nothing* architecture paradigm [15, 16].

To achieve its goals, the e-sensing project proposed three work packages (WP) [1]:

1. Databases: This WP is about researching and developing array databases to store large Earth Observation data sets. It also develops work flows and methods for efficient storage, access and processing of large data sets.

2. Data analysis: This WP is about researching and developing spatiotemporal techniques for extracting change information on large Earth Observation data sets. This is relevant, for example, for forestry applications. This WP includes finding novel applications of remote sensing time series, and combining time series with multitemporal image processing.

3. Use case development: This WP comprises the development of applications for forestry and agriculture management where large Earth Observation data sets are useful. The use cases derived from these applications will validate the methods and data developed by the other work packages.

The first work package, databases, is composed of the tasks *Building and deployment of big Earth observation databases to support data analysis and use cases* and also *Extend*

SciDB for geographical data handling. The first one is concerned with the data required to perform analysis and the second one deals with the semantics and interoperability of spatial data. Together, this two tasks provide the foundations for the remaining work packages as they provide the e-sensing platform users with data sets and operations required by Earth Observation scientists. This report is concerned with the former task, which is is split in three parts:

1. Database building. This task consist on loading satellite images to an array database.
2. Radiometric correction module. Radiometric correction allows the comparison of satellites images. As satellite images differ on spatiotemporal coverage, radiometric correction removes unwanted influences such as atmospheric effect (such as haze) or sensor errors.
3. Geometric correction module. Geometric correction is concerned with adjusting images in such a way that features of interest overlap.

The sub-task *database building* includes the following data sets:

- The MODIS MOD09Q1 and MOD13Q1 images at 250 meter and weekly resolutions with temporal extent from 2000 to 2014 and the spatial extent of South America. This data set has a size of 15 Terabytes.
- Part of INPE's LANDSAT-5 data collection. Its temporal extent goes from 1984 to 2012, the spatial extent covers Brazil with temporal resolution of 6 coverages a year. This data set has a size of 30 Terabytes.
- A data selection of satellites SENTINEL-2A, CBERS-4 and LANDSAT-8. The size of this data set is 10 Terabytes.

2 Achievements of the period

2.1 Database building

This sub-task can be divided into the database scheme, the tools to upload data, and the data uploaded. The details are below.

2.1.1 Database schema

SciDB is a distributed database based on the array data model. It is distributed because SciDB executes several instances of itself across several servers. One of these instances plays the coordinator role and it is responsible for mediating client communications and for orchestrating query executions. The other instances — called workers — process the queries. SciDB was designed following a share-nothing architecture on which each server is responsible for its own resources (i.e. memory and processors). SciDB splits arrays into smaller parts, called chunks, that are distributed among all the servers in the database [15, 16].

SciDB users control the arrays through schemes. Schemes are defined in terms of dimensions, chunk sizes (one per dimension), overlap parameters (one per dimension), and attributes. The dimensions are the coordinate system used to locate data while the chunk sizes shape the units of storage; finally, the overlap parameters state the amount of data shared among chunks [16].

By making an abstraction of the satellite images taken by the same sensor, we designed a generic storing schema. This allowed us to develop tools for loading data to SciDB, as all the images fit the same schema.

- **Dimensions**. Each arrays has 3 dimensions: Two for space and one for time that are respectively called (*col_ id*, *row_ id*), and (*time_ id*). The spatial dimensions are an enumeration of pixels starting at *(0,0)* with the top left pixel of the top left tile. From there, the dimension value increments continuously to the right and down until the down right pixel of the down right tile is reached. Regarding the temporal dimension, it starts at 0 and it increases by one along the temporal resolution of the images. For example, the MOD09Q1 temporal resolution is one image a week, so the dimension starts at *0* for the image of the first week of January of 2000.

- **Chunk Size**. The chunk sizes for data loading are *col_id=75*, *row_id=75*, and *time_id=400*. The chunk size parameter impacts performance and it can (must be) adapted according to the kind of application. For example, time-series processing requires a larger chunk size along the time dimension than spatial analysis applications.

- **Overlap**. The chunk overlap for data loading is zero. For Earth Observation applications, an overlap between zero and five is considered appropriate on the spatial dimensions. On the other hand, for the temporal dimension, the overlap can range from one (for example, in *Markov-Chain* applications), to the whole time series (for example, when using *Dynamic Time Wrapping*) algorithm. However, since the raw data (the images) do not overlap, we usually set this parameter to zero.

- **Attributes**. The attributes in the schema vary according to each type of image. As mentioned before, the images taken by the same satellite-sensor share the same attributes. Organization such as NASA distribute their post-processed images as products, on which the organize the image attributes according to specific applications. For example MOD13Q1 is specifically designed for land cover applications [17].

2.1.2 Database loading tools

The *modis2scidb-loader* are the tools for loading satellite images to SciDB. The *modis2scidb-loader* are python scripts which orchestrate the Extraction, Transformation, Loading (ETL) process of data as follows:

- The extract component reads satellite image formats and makes data available for the transform component. This component relies on the Geospatial Data Abstraction Library (GDAL). GDAL allows applications to read several satellite image formats.

- The transform component arranges the image data into an structure that SciDB can read. SciDB reads many formats, however, this component specifically writes to the SciDB binary format because it produces smaller files, hence, reducing the amount of storage required.

- The load component sends instructions to SciDB to load the binary data produced by the component above. This component constantly monitors a directory, searching for SciDB binary files. As soon as a new files are added, this component sends instructions to load the data to the server.

These tools take as input image files, database parameters, and an array name. Their output is the data inside a SciDB array. The extract and transform run independently of the loading component; this provides the opportunity to run these components in parallel. These tools are documented and available through the e-sensing code repository at Github [1].

2.1.3 Uploaded data

- The MODIS products MOD09Q1 and MOD13Q1 are already uploaded. This year they were kept up to date with the new images released.

- Data from Tropical Rainfall Measuring Mission data has been also updated to the database. This data set is composed of satellite estimation of rainfall in the tropics. The data has a spatial resolution of 0.25 degrees of latitude and longitude and temporal resolution of one month. The data set extent goes from 1998 to 2006 and it covers the Amazon tropical forest.

- The Landsat products are currently under re-processing by the United States Geological Survey (USGS). According to the USGS website [2], the Landsat images are being re-processed and re-organized into collections, following a distribution model similar to MODIS. These changes include the algorithm used on the raw data, the image file identifiers, the metadata and the inclusion of a new quality band. The USGS reports these changes are meant to enable Landsat data for time-series analysis and the reprocessing prioritizes the most recent images covering the United States. The whole re-pocessing will take until April 2017. For this reason, loading Landsat data to SciDB hadn't been done yet.

- The data selection of satellites SENTINEL-2A, CBERS-4 and LANDSAT-8 haven't been loaded yet.

[1]Extract, Transform and Load for SciDB `https://github.com/e-sensing/scietl`
[2]Landsat collections`https://landsat.usgs.gov/`

2.2 Radiometric and Geometric correction modules

The radiometric module is still under development. However, we have successfully tested implementations of some radiometric corrections using SciDB's Array Functional Language (AFL). Sets of AFL operators can be grouped into a macro, then this macro is exposed as an additional AFL function. Macros can be grouped in SciDB as modules. This SciDB extension mechanism resembles the store procedures that are commonly found in Relational Database Management Systems (RDMS).

The radiometric and geometric corrections of satellite images are fall into Tomlin's classification as local or focal map algebra functions. Tomlin's map algebra organizes the functions that can be applied to images into three categories: local, zonal, and focal. Local operations compute a values for every pixel as a function of the existing pixel value. Zonal operations compute new values for every pixel as a function of the pixel values itself and their neighbours; these neighbours are determined by an arbitrary neighbourhood function. Focal operations compute new values for every location as a function of the pixels' values, distances, or directions of neighboring pixels. In the focal operations, the evaluated pixel and its neighbors fit square shapes unlike the zonal operations where the shape of the zone is arbitrary [18].

The radiometric corrections used in remote sensing match the local and focal functions. SciDB AFL includes operators able to do similar as Tomlin's map algebra. For example, the AFL *apply* calculates the local functions while *window* and *variable_window* match the focal functions.

The geometric correction module development hadn't started yet. The geometric corrections can be though of transformation of 2D plane coordinates, specifically the simmilarity, affine, and polynomial transformation [19, Chapter 5].

However, the test of the radiometric module provide some insights on the implementation of the geometric correction module. This module can be decomposed into three steps. First, computing the coordinates of the pixels, then apply a 2D coordinate transformation, then apply a re-sampling operation, and finally use the AFL operator *redimension* to obtain a geometrically corrected array.

2.3 Impact assessment

This database schema is a guide for the development of software tools to create, read, update, and delete data from the database. Besides, it is the foundation of a spatio-temporal data type which can be added to the existing geographic information system formats vector, raster, and triangulated irregular networks. Additionally, it play a role in the implementation of the Fields as a Generic Data Type for Big Spatial Data. Such a generic data type would enable the use of spatiotemporal arrays as source of other types such as coverage, time series, and trajectories [20].

The database loading tools enable the e-sensing project to build spatiotemporal arrays made of satellites images taken by different sensors. The results can be seen as the uploaded data allowed the publications of several articles (see Chapter 3).

3 Participation in scientific events

During the present period, the author of this report took part in the publications listed below. In general, his role consisted on preparing the data sets, setting up the database (SciDB), preparing the manuscript sections regarding the data.

- "Big earth observation data analytics" [1] published in the proceedings of the 5th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data [21].
- "Using dynamic geospatial ontologies to support information extraction from big Earth Events as key concepts for describing land use change"published in the proceedings of the ninth international conference on geographic information (GIScience 2016) [22].
- "Spatio-temporal change detection from multidimensional arrays: Detecting deforestation from MODIS time series"published in the ISPRS Journal of Photogrammetry and Remote Sensing [23]
- "A Time-Weighted Dynamic Time Warping Method for Land-Use and Land-Cover Mapping"published in the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing [24]

[1] The author of this report is mentioned in the "additional authors"section of the paper.

ANNEX 3.1

RELATÓRIO SUCINTO DO BOLSISTA POS-DOC

Rodrigo Anzolin Begotti

Subprojeto "Uso de séries temporais de imagens de satélites para monitorar Desmatamento e Degradação na floresta amazônica"

(processo FAPESP n° 16/16968-2)

O bolsista de pós-doutorado Rodrigo Anzolin Begotti, vinculado a este projeto por meio do subprojeto "Uso de séries temporais de imagens de satélites para monitorar Desmatamento e Degradação na floresta amazônica" (processo FAPESP n° 16/16968-2), iniciou suas atividades no dia 1 de outubro de 2016. Nesse período, o bolsista realizou a qualificação de parte do acervo de fotografias da fototeca do INPE (www.obt.inpe.br/fototeca) e do projeto "Videografia Aérea da Amazônia" (www.dpi.inpe.br/geoma/videografia) realizado no ano de 2006 pela mesma instituição. Ambos os acervos consistem de bancos de dados geográficos nos quais as informações de cada foto como missão, data, coordenadas geográficas, município e estado são armazenadas. Durante os três primeiros meses de vigência da bolsa, foram qualificadas 7013 fotos utilizando a nomenclatura do projeto PRODES. A qualificação das fotos se enquadra no objetivo de avaliar o desempenho das técnicas de classificação do desmatamento e degradação florestal desenvolvidas nesse projeto principal, inicialmente utilizando séries temporais de imagens MODIS.

Foram estabelecidos alguns critérios qualitativos para a qualificação do acervo de fotos. Somente as fotos das missões do INPE que realizaram sobrevoo foram utilizadas. Aquelas fotos cujo ângulo de captura foi menor a aproximadamente 60° em relação ao solo não foram qualificadas. Além disso, áreas fotografadas nas quais o centro da foto apresenta heterogeneidade de usos/cobertura da terra também não foram qualificadas. Dessa forma, de todas as fotos qualificadas, 4546 ou aproximadamente 12% do total pertencem ao acervo da Videografia. Do acervo da fototeca, foram qualificadas 2467 fotos ou 9,8% do total.

Durante os meses subsequentes, particularmente janeiro e fevereiro de 2017, o bolsista está auxiliando o grupo de desenvolvimento do projeto na coleta de amostras de usos da terra em escala temporal para o aprimoramento dos algoritmos de classificação. Nos próximos meses, o bolsista começará a trabalhar na aplicação desses algoritmos para a detecção de desmatamento e degradação florestal.

ANNEX 4.1

Instituto Nacional de Pesquisas Espaciais

Programa de Pós Graduação em Computação Aplicada

Relatório científico 01

FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo

Pesquisa de doutorado:

Métodos de análise de dados espaço-temporais

Projeto

Esensing: Big Earth observation data analytics for land use and land cover change information

FAPESP Research Program in e-science - Grant 2014/08398-6

Bolsista – Rennan de Freitas Bezerra Marujo

Orientadora – Prof. Leila Maria Garcia Fonseca

Processo n. 2016/08719-2

São José dos Campos – SP

Dezembro 2016

1 Introdução

Este é o primeiro relatório referente à bolsa de Doutorado fluxo contínuo - processo número 2016/08719-2 - outorgada à Rennan de Freitas Bezerra Marujo, aluno regular do Programa de Pós-Graduação em Computação Aplicada do Instituto Nacional de Pesquisas Espaciais. A bolsa financia a pesquisa “Métodos de análise de dados espaço-temporais” desenvolvida no projeto e-sensing - processo 2014/08398-6 - sob a orientação da Profa. Dra. Leila Maria Garcia Fonseca.

A Terra está em constante mudança, sendo a caracterização e mapeamento da cobertura terrestre essenciais para planejar e gerenciar seus recursos naturais. Entender os processos ativos, como o desmatamento, a expansão urbana e os fenômenos naturais, é vital para a preservação dos ecossistemas, dessa forma, é importante desenvolver ferramentas capazes de detectar tais variações (KUENZER et al., 2015).

O processo de detecção de mudanças é uma tarefa difícil de realizar, sendo comumente utilizado nesta operação sensores orbitais multi-espectrais (COPPIN et al., 2004). Sensores orbitais de alta resolução espacial captam informação da superfície da Terra com mais detalhes do que com sensores de baixa resolução espacial. Entretanto, há um compromisso entre as características de resoluções temporal, espacial e radiométrica que podem limitar o desempenho do sensor em algumas aplicações (LEFSKY, COHEN, 2003). Para preencher a lacuna entre alta resolução espacial e alta resolução temporal, foram projetados sensores de resolução espacial média (10 a 50 metros) (EHLERS et al., 2002).

O INPE foi pioneiro na provisão livre de imagens orbitais de média resolução espacial, liberando imagens do segundo Satélite Sino-Brasileiro de Recursos Terrestres (CBERS-2) gratuitamente (BANSKOTA et al., 2014). A adoção desta política encorajou o Serviço Geológico dos Estados Unidos (United States Geological Survey - USGS) a disponibilizar os dados Landsat em 2008 (WOODCOCK et al., 2008, BANSKOTA et al., 2014), o que resultou numa maior quantidade de acessos e utilização destas imagens (WULDER et al., 2012).

Os métodos para a detecção de mudança normalmente utilizam séries temporais curtas, variando de duas a cinco imagens, não utilizando o potencial completo das séries históricas (COPPIN et al., 2004). Neste contexto, as séries temporais podem fornecer observações de padrões, não encontradas em observações de data única, como tendências, periodicidades e modelos de previsão (EHLERS, 2009).

Séries temporais de imagens orbitais são o único conjunto contínuo e consistente de informações sobre a Terra (MAUS et al. , 2016b), integrando a informação espectral e espacial com a componente temporal, proporcionando informação rica para detalhar as variações espaciais ao longo do tempo (PETITJEAN et al., 2012). No entanto, a ausência de imagens de boa qualidade devido a presença de nuvens, baixa resolução temporal, bem como defeitos do sensor (por exemplo, as lacunas em imagens Landsat-7) exigem a sua correção e em muitas aplicações há a necessidade do uso de mais de um sensor (LEFSKY; COHEN, 2003;SHEN et al., 2016).

Atualmente, devido ao incremento na quantidade e disponibilidade das plataformas de sensoriamento remoto (WULDER et al., 2012), as abordagens multi-sensores tornaram-se mais promissoras, uma vez que diferentes sensores podem melhorar o mapeamento e o monitoramento das variáveis da vegetação ao longo do tempo (MOUSIVAND et al., 2015).

A abordagem multi-sensor para imagens ópticas pode contribuir na produção de produtos unificados, úteis nas várias aplicações de sensoriamento remoto (SAMAIN et al., 2006). Apesar disso, ainda são necessárias técnicas de processamento de imagens para unificar esses dados de forma que eles possam ser integrados em uma mesma base de dados prontos para análise (EHLERS, 2009), pois estes dados possuem diferentes resoluções espaciais, espectrais, temporais e angulares (MOUSIVAND et al., 2015).

Neste contexto, este relatório descreve as atividades acadêmicas e de pesquisa realizadas pelo bolsista desde março de 2016, quando se deu o ingresso no programa de pós-graduação, que envolvem conclusão das disciplinas, participação em eventos e redação de 4 artigos científicos. Neste relatório também são mostrados o resumo do projeto de pesquisa, mantido do plano de projeto da primeira etapa em Março de 2016 e por fim um plano de trabalho para as próximas etapas da pesquisa.

2 Resumo do projeto de pesquisa

2.1. Resumo

O Objetivo da pesquisa é desenvolver algoritmos de análise espaço-temporal para extrair informações de grandes bancos de imagens de observação da Terra. Este trabalho foca no desenvolvimento de métodos de análise espaço-temporal das mudanças de uso e cobertura da terra em grandes conjuntos de dados. Para isto serão desenvolvidos (1) técnicas de unificação dos dados obtidos de sensores diferentes que envolvem várias etapas: calibração radiométrica e geométrica, analisar os diferentes ângulos de aquisição, cobertura de nuvem, compatibilização das bandas espectrais dos sensores, etc; (2) geração das séries temporais; (3) realizar a análise das séries temporais aplicadas a detecção de mudanças no uso e cobertura da Terra. Em relação ao plano de trabalho do projeto de pesquisa, durante este período, foram desenvolvidas técnicas de unificação dos dados dos satélites Landsat-7 e CBERS-4. Além disso, foram estudados algumas técnicas de análise de séries temporais (BFAST, TWDTW, STARFM) que serão avaliadas no projeto.

Informamos que o desenvolvimento do algoritmo “Segmentação espaço-temporal”, que estava proposto no plano de trabalho, será desenvolvido pelo aluno de doutorado Wanderson Costa, também membro do projeto e-sensing, de modo que o trabalho de ambos é complementar. Deste modo o bolsista irá desenvolver métodos que combinam a abordagem multi-sensor com análise de séries temporais.

2.2. Objetivo

Desenvolver algoritmos de análise espaço-temporal para extrair informações de grandes bancos de imagens de observação da Terra.

2.3. Metas da tese de doutorado:

- Conceber, implementar e validar métodos de detecção de mudanças de uso e cobertura da terra em grandes bancos de dados com séries temporais extraídas de imagens de sensoriamento remoto multi-sensores;

- Publicar dois artigos em congresso internacionais e dois artigos em revista científicas.

2.4. **Cronograma proposto no plano inicial (Março 2016)**

Cronograma

Tarefa	1º Ano		2º Ano		3º Ano	
Métodos de análise de series temporais de imagens de sensoriamento remoto						
Segmentação e classificação espaçotemporal						
Artigos científicos						
Defesa da tese						

Os objetivos traçados para o primeiro ano de pesquisa foram alcançados, de modo que a pesquisa em métodos de análise de séries temporais de imagens de sensoriamento remoto vem sendo realizada e é parte da qualificação do bolsista.

No primeiro período, o autor submeteu um artigo científico para o Simpósio Brasileiro de Sensoriamento Remoto, de título: “CBERS-4/MUX automatic detection of clouds and cloud shadows using decision trees”, cuja resposta de aceite ou não será dado em janeiro. Estamos preparando um novo artigo para submissão na revista Journal of Computational Interdisciplinary Sciences (JCIS) de título “Raster Data Processing with TerraLib for Lua: na application to fill Landsat-7 SLC-off gaps”.

Em colaboração com o estudante de doutorado, Hugo Bendini estudante, foi submetido ao Simpósio Brasileiro de Sensoriamento Remoto o artigo de título: “Evaluation of smoothing methods on Landsat-8 EVI time series for crop classification based on phenological parameters”. Também juntamente com este aluno, foi submetido e aceito no Simpósio Brasileiro de Geoinformática (Geoinfo) o artigo de título: “Assessment of a Multi-Sensor Approach for Noise Removal on Landsat-8 OLI Time Series Using CBERS-4 MUX Data to Improve Crop Classification Based on Phenological Features”.

2.5. **Cronograma alterado proposto para restante da pesquisa:**

Cronograma

Tarefa	1ºAno		2ºAno		3ºAno	
Estudar métodos de análise de séries temporais de sensoriamento remoto	X	X				
Desenvolvimento métodos para séries temporais multi-sensor						
Artigos científicos		X				
Defesa de tese						

Deste modo, no decorrer do próximo ano o bolsista continuará sua pesquisa em análise de séries temporais focando na abordagem multi-sensor.

3 Resumo das atividades desenvolvidas

3.1. Disciplinas

O bolsista cursou 25 créditos em disciplinas, de um total necessário de 24 créditos, incluindo disciplinas obrigatórias, no período de Março a Dezembro de 2016. As disciplinas cursadas foram: Seminários em Computação Aplicada II, Computação Aplicada I, Matemática Computacional I, Inteligência Artificial, Fundamentos de Programação Estruturada, Tópicos em Observação da Terra, Processamento Digital de Imagens, Banco de Dados Geográficos, Princípios e Aplicações de Mineração de Dados e Tópicos Avançados em Processamento de Imagens.

3.2. Pesquisa

No inicio do projeto o bolsista fez um levantamento sobre os satélites de média resolução espacial em atividade e os catalogou conforme a largura de suas bandas e resolução temporal. Para construir uma série temporal multi-sensores, algumas análises foram feitas. Foi decidido utilizar inicialmente uma série Landsat-7/ETM+ com CBERS-4/MUX por terem características espectrais similares. Futuramente, essa série será ampliada para incluir dados do Landsat-8/OLI, como já vem sendo trabalhado por Holden et al., 2016, em relação ao Landsat-7 e Landsat-8. Essa série será integrada com a série Landsat devido a esta prover a maior série histórica de imagens orbitais desde 1972 (COHEN & GOWARDS, 2004). Por fim deseja-se abstrair essa integração para múltiplos sensores.

3.3. Submissão de artigos:

Baseado nos estudos realizados até o momento, dois artigos científicos foram produzidos pelo bolsista: “Raster Data Processing with TerraLib for Lua: an application to fill Landsat-7 SLC-off gaps” e “CBERS-4/MUX automatic detection of clouds and cloud shadows using decision trees”. O primeiro artigo está sendo revisado para ser submetido para um revista cientifica. O segundo artigo foi submetido para o Simpósio Brasileiro de Sensoriamento Remoto (SBSR) e terá notificação sobre o aceite em 16 de janeiro de 2017.

O bolsista participou como co-autor de dois artigos com o aluno de doutorado Hugo Bendini, sendo um aceito para publicação no Simpósio Brasileiro de

Geoinformática (Geoinfo) e o outro submetido aguardando aceite no Simpósio Brasileiro de Sensoriamento Remoto.

3.4. **Participação em eventos**

Durante o período relatado o bolsista participou como ouvinte do 16º Workshop de Computação Aplicada do INPE (WORCAP 2016) e da conferencia internacional 29th Conference on Graphics, Patterns and Images (SIBGRAPI).

3.5. **Bibliografia**

COHEN, W. B.; GOWARD, S. N. Landsats Role in Ecological Appli-252cations of Remote Sensing. BioScience, v.54, n.6, p.535546. 2004.

Holden, E. C., Curtis, E. W., (in press) 2016. An analysis of Landsat 7 and Landsat 8 underflight data and the implications for time series investigations. Remote Sensing of Environment.

remote sensing

Article

Cloud Cover Assessment for Operational Crop Monitoring Systems in Tropical Areas

Isaque Daniel Rocha Eberhardt [1,*,†], Bruno Schultz [2,†], Rodrigo Rizzi [3], Ieda Del'Arco Sanches [2], Antonio Roberto Formaggio [2], Clement Atzberger [4], Marcio Pupin Mello [5], Markus Immitzer [4], Kleber Trabaquini [6], William Foschiera [2] and Alfredo José Barreto Luiz [7]

1 Universidade de Brasília, Campus Universitário Darcy Ribeiro, Pós-graduação em Transportes, Anexo SG-12, 1° andar, CEP 70910-900 Brasilia, Brazil
2 Instituto Nacional de Pesquisas Espaciais (INPE), Divisão de Sensoriamento Remoto (DSR), Avenida dos Astronautas 1.758, 12227-010 São José dos Campos, Brazil; schultz@dsr.inpe.br (B.S); ieda@dsr.inpe.br (I.D.A.S); formag@dsr.inpe.br (A.R.F.); wfoschiera@dsr.inpe.br (W.F.)
3 Universidade Federal de Pelotas—UFPel, Campus Universitário S/N, Caixa postal 354, 96001-970 Pelotas-RS, Brazil; rodrigo.rizzi@ufpel.edu.br
4 Institute of Surveying, Remote Sensing and Land Information (IVFL), University of Natural Resources and Life Sciences, Vienna (BOKU), Peter Jordan Strasse 82, 1190 Vienna, Austria; clement.atzberger@boku.ac.at (C.A.); markus.immitzer@boku.ac.at (M.I.)
5 The Boeing Company, Boeing Research and Technology Brazil, 12247-016 São José dos Campos-SP, Brazil; marcio.p.mello@boeing.com
6 Agricultural Research and Rural Extension Company of Santa Catarina–Epagri. Santa Catarina State Information Center on Environmental and Hydrometeorological Resource-CIRAM. Rod. Admar Gonzaga, Itacorubi 1.347, Caixa Postal 502, 88034-901 Florianópolis-SC, Brazil; klebertrabaquini@epagri.sc.gov.br
7 Embrapa Environment, Caixa Postal 69, 13820-000 Jaguariúna-SP, Brazil; alfredo.luiz@embrapa.br
* Correspondence: isaquedanielre@hotmail.com; Tel.: +55-61-9901-56568
† These authors contributed equally to this work.

Academic Editors: Richard Gloaguen and Prasad S. Thenkabail
Received: 5 November 2015; Accepted: 1 March 2016; Published: 8 March 2016

Abstract: The potential of optical remote sensing data to identify, map and monitor croplands is well recognized. However, clouds strongly limit the usefulness of optical imagery for these applications. This paper aims at assessing cloud cover conditions over four states in the tropical and sub-tropical Center-South region of Brazil to guide the development of an appropriate agricultural monitoring system based on Landsat-like imagery. Cloudiness was assessed during overlapping four months periods to match the typical length of crop cycles in the study area. The percentage of clear sky occurrence was computed from the 1 km resolution MODIS Cloud Mask product (MOD35) considering 14 years of data between July 2000 and June 2014. Results showed high seasonality of cloud occurrence within the crop year with strong variations across the study area. The maximum seasonality was observed for the two states in the northern part of the study area (*i.e.*, the ones closer to the Equator line), which also presented the lowest averaged values (15%) of clear sky occurrence during the main (summer) cropping period (November to February). In these locations, optical data faces severe constraints for mapping summer crops. On the other hand, relatively favorable conditions were found in the southern part of the study region. In the South, clear sky values of around 45% were found and no significant clear sky seasonality was observed. Results underpin the challenges to implement an operational crop monitoring system based solely on optical remote sensing imagery in tropical and sub-tropical regions, in particular if short-cycle crops have to be monitored during the cloudy summer months. To cope with cloudiness issues, we recommend the use of new systems with higher repetition rates such as Sentinel-2. For local studies, Unmanned Aircraft Vehicles (UAVs) might be used to augment the observing capability. Multi-sensor approaches combining optical and microwave data can be another option. In cases where wall-to-wall maps are not mandatory, statistical sampling approaches might also be a suitable alternative for obtaining useful crop area information.

ISPRS Journal of Photogrammetry and Remote Sensing xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Spatio-temporal change detection from multidimensional arrays: Detecting deforestation from MODIS time series

Meng Lu [a,*], Edzer Pebesma [a], Alber Sanchez [a,b], Jan Verbesselt [c]

[a] Institute for Geoinformatics, Westfälische Wilhelms-Universitt Münster (WWU), Heisenbergstraße 2, 48149 Münster, Germany
[b] National Institute for Space Research (INPE), Av. dos Astronautas, 1758, 12227-007 São José dos Campos, Brazil
[c] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Droevendaalsesteeg 3, Wageningen 6708 PB, The Netherlands

a r t i c l e i n f o

Article history:
Received 8 June 2015
Received in revised form 19 January 2016
Accepted 9 March 2016
Available online xxxx

Keywords:
BFAST
Time series analysis
Spatial correlation
Temporal correlation
Array database
Spatio-temporal change modeling

a b s t r a c t

Growing availability of long-term satellite imagery enables change modeling with advanced spatio-temporal statistical methods. Multidimensional arrays naturally match the structure of spatio-temporal satellite data and can provide a clean modeling process for complex spatio-temporal analysis over large datasets. Our study case illustrates the detection of breakpoints in MODIS imagery time series for land cover change in the Brazilian Amazon using the BFAST (Breaks For Additive Season and Trend) change detection framework. BFAST includes an Empirical Fluctuation Process (EFP) to alarm the change and a change point time locating process. We extend the EFP to account for the spatial autocorrelation between spatial neighbors and assess the effects of spatial correlation when applying BFAST on satellite image time series. In addition, we evaluate how sensitive EFP is to the assumption that its time series residuals are temporally uncorrelated, by modeling it as an autoregressive process. We use arrays as a unified data structure for the modeling process, R to execute the analysis, and an array database management system to scale computation. Our results point to BFAST as a robust approach against mild temporal and spatial correlation, to the use of arrays to ease the modeling process of spatio-temporal change, and towards communicable and scalable analysis.

1. Introduction

Advanced earth observation satellite sensors provide remote sensing products that are rich in spatial, temporal, and spectral information. Open access policies of space agencies and the progress of remote sensing technologies make these products more accessible, which enables a wide range of novel applications, such as near real-time global change monitoring. This, however, calls for efficient handling and scalable processing of the massive amounts of available data. Major challenges include big data management, multidimensional data information extraction, and complex large-scale spatio-temporal change modeling procedures implementation and result visualization. These challenges call for novel data management and analytics tools and advanced spatio-temporal statistical algorithms.

Typical remote sensing satellite images are regularly discretised in space and time, and can naturally be represented as multidimensional arrays. The array data structure facilitates change modeling in many ways. Firstly, the array data structure allows a clean data processing procedure which simplifies data preparation, and avoid data structure conversions during the analysis. Wickham (2014) calls the unified data preparing process to "tidy data", and suggests restructuring all datasets into single, long tables. Since most earth observation data (i.e. earth information collected by remote sensing technologies) come as time series of multispectral images, and structuring such datasets into arrays is the more natural approach for data storage, analysis and visualization. In addition, the array data structure allows flexible application of spatio-temporal statistical algorithms (Zscheischler et al., 2013) and other information extraction methodologies (Mello et al., 2013), which was already exploited in the on-line analytical processing (OLAP) approach to analyze business data (Chaudhuri and Dayal, 1997; Viswanathan and Schneider, 2011). Finally, the array data structure facilitates parallelizing of the modeling process (Stonebraker et al., 2013). Array Data Management and Analytics Software (DMAS), which stores and operates on data as multidimensional arrays, can thus be used to scale the process and resolve the difficulties of large memory consumption and computational bottlenecks usually found in non-parallelized systems.

* Corresponding author.
E-mail address: meng.lu@uni-muenster.de (M. Lu).

http://dx.doi.org/10.1016/j.isprsjprs.2016.03.007

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

A Time-Weighted Dynamic Time Warping Method for Land-Use and Land-Cover Mapping

Victor Maus, *Member, IEEE*, Gilberto Câmara, Ricardo Cartaxo, Alber Sanchez, Fernando M. Ramos, and Gilberto R. de Queiroz

***Abstract*—This paper presents a time-weighted version of the dynamic time warping (DTW) method for land-use and land-cover classification using remote sensing image time series. Methods based on DTW have achieved significant results in time-series data mining. The original DTW method works well for shape matching, but is not suited for remote sensing time-series classification. It disregards the temporal range when finding the best alignment between two time series. Since each land-cover class has a specific phenological cycle, a good time-series land-cover classifier needs to balance between shape matching and temporal alignment. To that end, we adjusted the original DTW method to include a temporal weight that accounts for seasonality of land-cover types. The resulting algorithm improves on previous methods for land-cover classification using DTW. In a case study in a tropical forest area, our proposed logistic time-weighted version achieves the best overall accuracy of 87.32%. The accuracy of a version with maximum time delay constraints is 84.66%. A time-warping method without time constraints has a 70.14% accuracy. To get good results with the proposed algorithm, the spatial and temporal resolutions of the data should capture the properties of the landscape. The pattern samples should also represent well the temporal variation of land cover.**

***Index Terms*—Dynamic programming, image sequence analysis, monitoring, pattern classification, time series.**

I. Introduction

THERE is a global increase in food and energy production from agriculture to keep 7.3 billion people. To support sustainable practices and find out about unsustainable uses of natural resources, good quality land-use and land-cover datasets are essential [1]. Earth observation satellites are the only source that provides a continuous and consistent set of information about the Earth's land and oceans. Since remote sensing satellites revisit the same place repeatedly, we can calibrate their images so that measures of the same place in different times are comparable. These observation can be organized in regular time intervals, so that each measure from sensor is mapped into a three-dimensional (3-D) array in space-time.

From a data analysis perspective, researchers then have access to space-time datasets. This has lead to much recent research on satellite image time-series analysis. Algorithms for analyzing image time series include methods for time-series reconstruction [2], detecting trend and seasonal changes [3]–[5], extracting seasonality information [6], land-cover mapping [7], detecting forest disturbance and recovery [8]–[10], crop classification [11]–[13], planted forest mapping [14], and crop expansion and intensification [15], [16].

Research on time-series data mining shows that methods based on dynamic time warping (DTW) have achieved significant results in many applications [17]–[19]. DTW works by comparing a temporal signature of a known event (e.g., a person's speech) to an unknown time series (e.g., a speech record of unknown origin) [17], [20]–[23]. The algorithm compares two time series and finds their optimal alignment, providing a dissimilarity measure as a result [23]. DTW provides a robust distance measure for comparing time series, even if they are irregularly sampled [13] or are out of phase in the time axis [24]. The large range of applications of digital time warping for time series analysis motivated our idea of using DTW for remote sensing applications.

The DTW method works well for shape matching, but is not suited *per se* for remote sensing time-series classification. It disregards the temporal range when finding the best alignment between two time series [23], [25]. Each land-cover class has a distinct phenological cycle that is relevant for space-time classification [26], [27]. Therefore, a good time-series land-cover classifier needs to balance between shape matching and temporal alignment. For example, although crops tend to vary their annual phenological cycles, these variations will not be extreme. Consider a set of samples of soybean whose cycles range from 90 to 120 days. A time series with similar shape but with much larger cycle is unlikely to come from a soybean crop. The standard DTW method warps time to match the two series. To avoid such mismatches, we introduce a time constraint that helps to distinguish between different types of land-use and land-cover classes.

Recent papers by [13] and [28] have used DTW for satellite image time-series classification. The method proposed in these papers sets a maximum time delay to avoid inconsistent temporal distortions based on the date of the satellite images. The time series is split in 1 year segments to match the agricultural phenological cycle in Europe. However, this temporal

Manuscript received July 08, 2015; revised November 06, 2015; accepted January 07, 2015. Part of this work was developed in the Young Scientists Summer Program at the International Institute for Systems Analysis, Laxenburg (Austria). This work was supported in part by the Institute for Geoinformatics, University of Münster (Germany), and in part by the Earth System Science Center, National Institute for Space Research (Brazil). The work of G. Camara was supported by CAPES under Grant 23038.0075692012-16, FAPESP e-science program under Grant 2014-08398-6, and CNPq under Grant 3121512014-4.

V. Maus and G. Câmara are with the National Institute for Space Research, 12227010 São José dos Campos, Brazil, and also with the Institute for Geoinformatics, University of Münster, 48149 Münster, Germany (e-mail: vwmaus1@gmail.com).

R. Cartaxo, A. Sanchez, F. M. Ramos, and G. R. de Queiroz are with the National Institute for Space Research, 12227010 São José dos Campos, Brazil.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2016.2517118 *remote sensing*

Article

Forest Fragmentation in the Lower Amazon Floodplain: Implications for Biodiversity and Ecosystem Service Provision to Riverine Populations

Vivian Renó [1,*], Evlyn Novo [1] and Maria Escada [2]

[1] Remote Sensing Division (DSR), Brazilian Institute for Space Research (INPE), Av. dos Astronautas 1758, São José dos Campos 12227-010, São Paulo, Brazil; evlyn.novo@inpe.br

[2] Image Processing Division (DPI), Brazilian Institute for Space Research (INPE), Av. dos Astronautas 1758, São José dos Campos 12227-010, São Paulo, Brazil; isabel.escada@inpe.br

* Correspondence: vivian.reno@inpe.br; Tel.: +55-12-3208-64-86

Academic Editors: Parth Sarathi Roy and Prasad S. Thenkabail
Received: 1 July 2016; Accepted: 17 October 2016; Published: 27 October 2016

Abstract: This article analyzes the process of forest fragmentation of a floodplain landscape of the Lower Amazon over a 30-year period and its implications for the biodiversity and the provision of ecosystem services to the riverine population. To this end, we created a multi-temporal forest cover map based on Landsat images, and then analyzed the fragmentation dynamics through landscape metrics. From the analyses of the landscape and bibliographic information, we made inferences regarding the potential impacts of fragmentation on the biodiversity of trees, birds, mammals and insects. Subsequently, we used data on the local populations' environmental perception to assess whether the inferred impacts on biodiversity are perceived by these populations and whether the ecosystem services related to the biodiversity of the addressed groups are compromised. The results show a 70% reduction of the forest habitat as well as important changes in the landscape structure that constitute a high degree of forest fragmentation. The perceived landscape alterations indicate that there is great potential for compromise of the biodiversity of trees, birds, mammals and insects. The field interviews corroborate the inferred impacts on biodiversity and indicate that the ecosystem services of the local communities have been compromised. More than 95% of the communities report a decreased variety and/or abundance of animal and plant species, 46% report a decrease in agricultural productivity, and 19% confirm a higher incidence of pests during the last 30 years. The present study provides evidence of an accelerated process of degradation of the floodplain forests of the Lower Amazon and indicate substantial compromise of the ecosystem services provision to the riverine population in recent decades, including reductions of food resources (animals and plants), fire wood, raw material and medicine, as well as lower agricultural productivity due to probable lack of pollination, impoverishment of the soil and an increase of pests.

Keywords: Amazon floodplain; forest fragmentation; remote sensing; biodiversity; ecosystem services; riverine population

---

1. Introduction

The human being depends on a wide variety of ecosystem services (ES) for its survival and well-being. Ecosystems are sources of indispensable resources such as food and water, and they also provide other important services such as pollination, erosion control, and air and water purification, among others [1,2]. Many of these services have supported the growing demands of modern societies. Currently, however, the majority of the ES on the planet are in decline due to anthropogenic disturbances [1,3].

Earth Interactions • Volume 20 (2016) •
Paper No. 17 • Page 1

Forest Degradation Associated with Logging Frontier Expansion in the Amazon: The BR-163 Region in Southwestern Pará, Brazil

T. F. Pinheiro[a]

Center for Earth System Science, National Institute for Space Research (INPE), São José dos Campos, Brazil

M. I. S. Escada

Center for Earth System Science, Department of Image Processing, National Institute for Space Research (INPE), São José dos Campos, Brazil

D. M. Valeriano

Amazonia Monitoring Program, Department of Remote Sensing, National Institute for Space Research (INPE), São José dos Campos, Brazil

P. Hostert

Geomatics Lab, and Department of Geography, Humboldt-Universität zu Berlin, Berlin, Germany

F. Gollnow and H. Müller

Department of Geography, Humboldt-Universität zu Berlin, Berlin, Germany

Received 19 February 2015; in final form 24 February 2016

---

[a] Corresponding author address: T. F. Pinheiro, Departamento de Ciências do Sistema Terrestre, Instituto Nacional de Pesquisas Espaciais, 1758 Av. Astronautas, CEP 12227-010, São José dos Campos, SP Brazil.

E-mail address: taise.pinheiro@inpe.br

DOI: 10.1175/EI-D-15-0016.1

Revista Brasileira de Cartografia (2016), Nº 68/4, Edição Especial Geoinformação e Análise Espacial: 717-731 Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto
ISSN: 1808-0936

MINERAÇÃO DE TRAJETÓRIAS DE MUDANÇA DE COBERTURA DA TERRA EM ESTUDOS DE DEGRADAÇÃO FLORESTAL

Land-Cover Change Trajectory Mining in Forest Degradation Studies

Marcio Azeredo[1], Antônio Miguel Vieira Monteiro[2], Maria Isabel Sobral Escada[2], Karine Reis Ferreira[2], Lúbia Vinhas[2] & Taise Farias Pinheiro[3]

[1]Diretoria de Serviço Geográfico do Exército – DSG
Quartel General do Exército - Bloco "F" - 2º Piso - SMU, Brasília – DF, 70630-901, Brasil
azeredo.marcio@eb.mil.br

[2]Instituto Nacional de Pesquisas Espaciais – INPE
Divisão de Processamento de Imagens – DPI
Av. dos Astronautas, 1.758, Jd. da Granja, São José dos Campos - SP, 12227-010, Brasil
{miguel,isabel,karine,lubia}@dpi.inpe.br

[3]Instituto Nacional de Pesquisas Espaciais – INPE
Centro de Ciênca do Sistema Terrestre – CCST
Av. dos Astronautas, 1.758, Jd. da Granja, São José dos Campos - SP, 12227-010, Brasil
taisefarias@hotmail.com

Recebido em 9 de Fevereiro, 2015/ Aceito em 7 de Agosto, 2015
Received on February 9, 2015/ Accepted on August 7, 2015

RESUMO

A cobertura da terra encontra-se em constante transformação devido aos diferentes fatores que influenciam a sua utilização e destinação. Essas alterações observadas ao longo do tempo definem o que chamamos de Trajetórias de Mudança de Cobertura da Terra. Tais trajetórias, uma vez caracterizadas, podem fornecer informações valiosas acerca de diferentes processos ambientais, como por exemplo, a degradação florestal. No entanto, existem poucas ferramentas disponíveis para a caracterização de trajetórias de mudança de cobertura da terra, pois a maior parte da literatura sobre trajetórias considera o conceito no contexto de estudo dos objetos móveis, com outra semântica associada. Com o objetivo de extrair e caracterizar trajetórias no contexto da mudança de cobertura da terra, este trabalho aproveita os conceitos já existentes sobre trajetórias e apresenta uma redefinição para alguns dos padrões de comportamento usualmente utilizados no contexto das trajetórias de objetos móveis. Para demonstrar suas possibilidades de utilização, foram conduzidos dois experimentos em uma base de dados de degradação florestal da região amazônica, onde os resultados permitiram identificar e caracterizar algumas das trajetórias presentes.

Palavras chaves: Trajetórias, Mudança de Cobertura, Padrões de Comportamento.

ABSTRACT

Land cover is changing as a consequence of different factors which influence its use and purpose. These observed changes over time and space define what we call Land-Cover Change Trajectories. Once characterized, such trajectories could provide valuable information on environmental issues as tropical forest degradation processes. However, there are few studies in the literature available tackling Land-Cover Change Trajectories Patterns detection and characterization.

Brazilian Journal of Cartography (2016), Nº 68/6, Special Issue GEOINFO 2015: 1113-1122 Brazilian Society of Cartography, Geodesy, Photgrammetry and Remote Sense
ISSN: 1808-0936

IMPROVEMENTS OF THE DIVIDE AND SEGMENT METHOD FOR PARALLEL IMAGE SEGMENTATION

Avanços do Método "Divide and Segment" para Segmentação de Imagens em Paralelo

Anderson Reis Soares, Thales Sehn Körting & Leila Maria Garcia Fonseca

Instituto Nacional de Pesquisa Espaciais – INPE
Divisão de Processamento de Imagens - DPI
Caixa Postal 515 – 12.245-970– São José dos Campos – SP – Brazil
{anderson.soares, thales.korting, leila.fonseca}@inpe.br

Received on December 5, 2015/ Accepted on February 27, 2016
Recebido em 5 de Dezembro, 2015/ Aceito em 27 de Fevereiro, 2016

ABSTRACT

Remote Sensing is an important source of data about the dynamics of Earth's land and oceans, but retrieve information from this technique, is a challenge. Segmentation is a traditional method in remote sensing, which have a high computational cost. An alternative to suppress this problem is use parallel approaches, which split the image into tiles, and segment each one individually. However, the divisions among tiles are not natural, which create inconsistent objects. In this work, we extended our previous work, which used non-crisp borders computed based on graph-theory. By applying this non-crisp line cut, we avoid the post-processing of neighboring regions, and therefore speed up the segmentation.

Key words: Image Segmentation, Parallel, Graph Theory.

RESUMO

Sensoriamento Remoto é uma importante fonte de dados sobre a dinâmica da superfície e dos oceanos. Contudo extrair informação e obter conhecimento através dessa técnica é um desafio. A segmentação é um método tradicional em Sensoriamento Remoto, que apresenta um grande custo computacional. Uma alternativa para suprimir esse problema é a utilização de abordagens que se valem do processamento em paralelo, dessa forma a imagem é dividida em blocos e cada um é segmentado individualmente. Entretanto, a forma para divisão dos blocos não é natural, resultando em objetos inconsistentes. Nesse artigo, são apresentados alguns avanços dos nossos trabalhos anteriores, por meio de bordas adaptativas obtidas através de métodos desenvolvidos para grafos. Aplicando as bordas adaptativas, foi possível evitar a criação de objetos inconsistentes e o uso de algoritmos de pós-processamento e, consequentemente, reduzir o tempo de processamento.

Palavras chave: Segmentação de Imagens, Processamento em Paralelo, Teoria dos Grafos.

Revista Brasileira de Cartografi a (2016) N0 68/1: 131-143
Sociedade Brasileira de Cartografi a, Geodésia, Fotogrametria e Sensoriamento Remoto ISSN: 1808-0936

CLASSIFICAÇÃO ORIENTADA A OBJETOS EM IMAGENS MULTITEMPORAIS LANDSAT APLICADA NA IDENTIFICAÇÃO DE CANA-DE-AÇÚCAR E SOJA

Object-Based Classification in Multi-Temporal Landsat Images Applied to Identify Sugarcane and Soybean

Bruno Schultz[1], Antonio Roberto Formaggio[1], Isaque Daniel Rocha Eberhardt[2], Ieda Del'Arco Sanches[1], Júlio Cesar de Oliveira[3] & Alfredo José Barreto Luiz[4]

[1] Instituto Nacional de Pesquisas Espaciais – INPE
Divisão de Sensoriamento Remoto – DSR
Caixa Postal 515 – 12227-010 – São José dos Campos, SP –Brasil
{schultz, formagg, ieda,}@dsr.inpe.br

[2] Universidade de Brasília – UNB
Campus Universitário Darcy Ribeiro, Pós-graduação em Transportes
Anexo SG-12, 1° andar, CEP 70910-900, Brasília, DF
isaquedanielre@hotmail.com

[3] Universidade Federal de Viçosa – UFV
Departamento de Engenharia Civil
CEP 36570-900, Viçosa, MG
oliveirajc@ufv.br

[4] Empresa Brasileira de Pesquisa Agropecuária, Brasil
EMBRAPA Meio Ambiente
Rod SP 340, km 127,5 Cx Postal 69 13820-000 Jaguariúna, SP
alfredo.luiz@embrapa.br

Recebido em 19 de Novembro, 2015/ Aceito em 06 de Fevereiro, 2015
Received on November 19, 2015/ Accepted on February 06, 2015

RESUMO

A presente pesquisa teve por objetivo avaliar a potencialidade de dados multitemporais Landsat para classificação de cana-de-açúcar e de soja, conjuntamente, quando realizada via Análise de Imagens Orientada a Objetos (OBIA/ Random Forest). Foi utilizado um segmentador multi-resolução (SM) para gerar os polígonos (objetos). Um conjunto de 500 segmentações foi criado pela variação dos parâmetros Fe (fator de escala), Fm (forma) e Cp (compacidade), e avaliado pelo Índice de Avaliação da Segmentação (IAVAS). Da segmentação que obteve menor IAVAS, foram extraídos os atributos espectrais das médias e desvios-padrão das bandas TM/Landsat-5 [setembro (S) e outubro (O) do ano 2000] e ETM+/Landsat-7 [fevereiro (F) e março (M) do ano 2001] dos objetos, e seus NDVIs. Estes atributos foram inseridos no algoritmo Random Forest (RF) e as exatidões das classificações foram testadas quanto ao uso dos seguintes conjuntos de datas: SOFM; SFM; OFM; SOF; FM; OF; SF; e F. O IAVAS definiu Fe (35), Fm (30) e Cp

Brazilian Journal of Cartography (2016), Nº 68/6, Special Issue GEOINFO 2015: 1097-1111=Brazilian Society of Cartography, Geodesy, Photgrammetry and Remote Sense
ISSN: 1808-0936

WATER BODY EXTRACTION FROM RAPIDEYE IMAGES: AN AUTOMATED METHODOLOGY BASED ON HUE COMPONENT OF COLOR TRANSFORMATION FROM RGB TO HSV MODEL

Extração de Corpos d'Água Utilizando Imagens RapidEye: Metodologia Automatizada com Base no Componente Matiz da Transformação de Cores RGB para o Modelo HSV

Laércio Massaru Namikawa, Thales Sehn Körting & Emiliano Ferreira Castejon

Brazil's National Institute for Space Research – INPE
Image Processing Division – DPI
Av. dos Astronautas, 1758 São José dos Campos, Brazil
{laercio, thales, castejon}@dpi.inpe.br

Received on December 17, 2015/ Accepted on February 27, 2016
Recebido em 17 de Dezembro, 2015/ Aceito em 27 de Fevereiro, 2016

ABSTRACT

Water management and flood studies are some fields in which a map with all water bodies in a region is useful, especially in scenarios of environmental changes due to anthropogenic factors. Various detection methods of water body surfaces in remotely sensed images are available, from simple methods having a lower accuracy to more sophisticated ones. The objective of this paper is to present a simple, yet accurate method to detect water bodies in RapidEye images. The motivation is the availability of country wide coverage of these images, which makes feasible the generation of a map of all water bodies detectable at that spatial resolution. Our solution is the use the color transformation from Red-Green-Blue to Hue-Saturation-Value and the minimum radiance from all RapidEye bands to classify water bodies in seven classes of water. The water classes are ranked based on the confidence of the classified pixels being water, which accommodates for the differences in illumination and scattering that are present in such a large coverage, composed by more than 15000 scenes. In addition, users of the generated water bodies map can reclassify based on their needs. The methodology was developed on two RapidEye scenes, covering the Jacareí and Foz do Iguaçu municipalities, in Brazil. Results indicate that the classification is better than the traditional ones, with the advantage of providing seven classes with confidence levels.

Keywords: RapidEye, Water Body Detection, RGB-HSV Color Transformation.

RESUMO

Gerenciamento de bacias hidrográficas e estudos sobre enchentes são exemplos de casos onde é útil a existência de um mapa contendo todos os corpos d'água de uma região, especialmente em cenários de mudanças ambientais abruptas, devido a fatores antropogênicos. Existem vários métodos de detecção de superfícies de corpos d'água utilizando imagens de sensoriamento remoto, desde os mais simples contendo baixa acurácia até os mais sofisticados. O objetivo deste trabalho é apresentar um método que combine uma taxa de acerto adequada a uma metodologia simples, para ser aplicado em imagens do satélite RapidEye. A motivação para a gerar este método foi a disponibilidade de uma cobertura completa do Brasil com imagens RapidEye. Esta base de imagens permite a descoberta de corpos d'água detectáveis através da resolução espacial do RapidEye. Para isso, deve-se aplicar a transformação de cores RGB (siglas em inglês

Brazilian Journal of Cartography (2016), Nº 68/6, Special Issue GEOINFO 2015: 1079-1086 Brazilian Society of Cartography, Geodesy, Photgrammetry and Remote Sense
ISSN: 1808-0936

TRENDS IN GEOINFORMATICS

Tendências em GeoInformática

Thales Sehn Körting, Karine Reis Ferreira, Lúbia Vinhas, Antônio Miguel Vieira Monteiro & Gilberto Ribeiro de Queiroz

National Institute for Space Research - INPE Brazil
DPI - Image Processing Division at the OBT - General Earth Observation Coordination
Av. dos Astronautas,1758 São José dos Campos, Brazil
{thales.korting, karine.ferreira, lubia.vinhas, miguel.monteiro, gilberto.queiroz}@inpe.br

ABSTRACT

After a thorough review of the past editions of the Brazilian Symposium on GeoInformatics (GEOINFO), from 1999 to 2015, which included more than 300 articles, we have identified some topics pointing to trends in the area of GeoInformatics for the next years of research. As expected, the topics *Data/Information*, *Geographic Information System (GIS)*, *Time* and *Space* were included as the most relevant terms. From this perspective, we understand that research in GeoInformatics for the next years will follow these trends, namely big data, spatiotemporal data analysis, geographical applications for mobile devices, new architectures of Spatial Data Infrastructure (SDI), and spatiotemporal visualisation methods guiding exploratory data analysis.

Keywords: Trends, Big Data, Data Analysis, Mobile Devices, Spatial Data Infrastructure, Visualization.

RESUMO

Após uma revisão minuciosa das edições passadas do GEOINFO (Simpósio Brasileiro de GeoInformática), entre 1999 e 2015, incluindo mais de 300 artigos, foram identificados alguns tópicos que podem apontar para tendências na área de GeoInformática para os próximos anos de pesquisa. Como esperado, termos como D*ados/Informações*, *Sistema de Informação Geográfica (SIG)*, T*empo* e E*spaço* se destacaram como mais relevantes. A partir desta perspectiva, este artigo propõe o que se entende como tendências de pesquisa na área de GeoInformática. Essas tendências incluem *big data*, análise de dados espaço-temporais, aplicativos geográficos para dispositivos móveis, novas arquiteturas de infraestrutura de dados espaciais, e métodos de visualização espaço-temporal para análise exploratória de dados.

Palavras chaves: Tendências, Análise de Dados, Dispositivos Móveis, Infraestrutura de Dados Espaciais, Visualização.

Brazilian Journal of Cartography (2016), Nº 68/6, Special Issue GEOINFO 2015: 1133-1142 Brazilian Society of Cartography, Geodesy, Photgrammetry and Remote Sense
ISSN: 1808-0936

COMBINING TIME SERIES FEATURES AND DATA MINING TO DETECT LAND COVER PATTERNS: A CASE STUDY IN NORTHERN MATO GROSSO STATE, BRAZIL

Uso de Séries Temporais e Mineração de Dados para Detectar Padrões de Cobertura da Terra: um Estudo de Caso no Norte do Mato Grosso, Brasil

Alana Kasahara Neves, Hugo do Nascimento Bendini,
Thales Sehn Körting & Leila Maria Garcia Fonseca

Instituto Nacional de Pesquisas Espaciais – INPE
Divisão de Processamento de Imagens - DPI
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
{alana.neves; thales.korting; leila.fonseca}@inpe.br
hnbendini@dpi.inpe.br

Received on December 5, 2015/ Accepted on February 27, 2016
Recebido em 5 de Dezembro, 2015/ Aceito em 27 de Fevereiro, 2016

ABSTRACT

One product of the MODIS sensor (Moderate Resolution Imaging Spectroradiometer) is the EVI2 (Two Band Enhanced Vegetation Index). It generates images of around 23 observations each year, that combined can be interpreted as time series. This work presents the results of using two types of features obtained from EVI2 time series: basic and polar features. Such features were employed in automatic classification for land cover mapping, and we compared the influence of using single pixel versus object-based observations. The features were used to generate classification models using the Random Forest algorithm. Classes of interest included Agricultural Area, Pasture and Forest. Results achieved accuracies up to 91,70% for the northern region of Mato Grosso state, Brazil.

Keywords: Time Series, Data Mining, Land Cover, MODIS.

RESUMO

Um dos produtos do sensor MODIS (Moderate Resolution Imaging Spectroradiometer) é o EVI2 (Two Band Enhanced Vegetation Index). O sensor é capaz de gerar cerca de 23 imagens da mesma região por ano, que combinadas podem ser interpretadas como uma série temporal. Esse trabalho apresenta os resultados do uso de dois tipos de atributos obtidos de séries temporais de EVI2: atributos básicos e polares. Tais atributos foram empregados na classificação automática para o mapeamento de cobertura da terra, e comparou-se a influência da utilização de pixels versus observações baseadas em objetos. Os atributos foram utilizados para gerar modelos de classificação usando o algoritmo Random Forest. As classes de interesse incluíram Agricultura Anual, Pastagem e Floresta. Os resultados atingiram taxas de acerto de até 91,70% para a região norte do Mato Grosso, Brasil.

Palavras chaves: Séries Temporais, Mineração de Dados, Cobertura da Terra, MODIS.

Big Earth Observation Data Analytics: Matching Requirements to System Architectures

Gilberto Camara
INPE - National Institute for Space Research
Av dos Astronautas 1758
Sao Jose dos Campos, Brazil
gilberto.camara@inpe.br

Luiz Fernando Assis
National Institute for Space Research
Av dos Astronautas 1758
Sao Jose dos Campos, Brazil
luizffga@dpi.inpe.br

Gilberto Ribeiro
National Institute for Space Research
Av dos Astronautas 1758
Sao Jose dos Campos, Brazil
gilberto.ribeiro@inpe.br

Karine Reis Ferreira
National Institute for Space Research
Av dos Astronautas 1758
Sao Jose dos Campos, Brazil
karine.ferreira@inpe.br

Eduardo Llapa
National Institute for Space Research
Av dos Astronautas 1758
Sao Jose dos Campos, Brazil
eduardo.llapa@inpe.br

Lubia Vinhas
National Institute for Space Research
Av dos Astronautas 1758
Sao Jose dos Campos, Brazil
lubia.vinhas@inpe.br

ABSTRACT

Earth observation satellites produce petabytes of geospatial data. To manage large data sets, researchers need stable and efficient solutions that support their analytical tasks. Since the technology for big data handling is evolving rapidly, researchers find it hard to keep up with the new developments. To lower this burden, we argue that researchers should not have to convert their algorithms to specialised environments. Imposing a new API to researchers is counterproductive and slows down progress on big data analytics. This paper assesses the cost of research-friendliness, in a case where the researcher has developed an algorithm in the **R** language and wants to use the same code for big data analytics. We take an algorithm for remote sensing time series analysis on compare it use on map/reduce and on array database architectures. While the performance of the algorithm for big data sets is similar, organising image data for processing in Hadoop is more complicated and time-consuming than handling images in SciDB. Therefore, the combination of the array database SciDB and the **R** language offers an adequate support for researchers working on big Earth observation data analytics.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; H.2.8 [**Database Applications**]: [Spatial databases and GIS, Scientific databases, Image databases]

Keywords

Earth Observation, Array Databases, Big Data Analytics

BigSpatial16, October31–November 03 2016, Burlingname, CA, USA

ISBN 978-1-4503-4581-1/16/10…$15.00

DOI: http://dx.doi.org/10.1145/3006386.3006393

1. INTRODUCTION

Earth observation (EO) satellites produce vast amounts of geospatial data. The Landsat archive holds over five million images of the Earth's surface, with about 1 PB of data. New satellites from Europe, USA, China, Brazil, and India generate yearly as much data as one Landsat satellite in a decade. Most space agencies have adopted an open data policy, making unprecedented amounts of satellite data available for research and operational use. This data deluge has brought about a major challenge for Geoinformatics research: *How to design and build technologies that allow the EO community to analyse big data sets?*

When scientists use big EO data they face the burden of organising thousands of files, downloaded from the space agencies archives. To manage such large data sets, researchers need stable and efficient solutions that support their analytical tasks. To choose a solution is hard because the technology for large data handling and analytics is evolving. Alternatives include MapReduce-based solutions such as Google Earth Engine [10], object-relational DBMS extensions such as Rasdaman [2] and distributed multidimensional array databases such as SciDB [26]. Since each of these architectures takes on a different approach, understanding the benefits and drawbacks of each one helps researchers choose what best fits their needs.

Given the diversity of options, researchers would gain from documented experience that helps them to assess how proposed big data architectures fit the needs of geospatial data analysis. Recent papers describe algorithms required for EO analysis [27] [21] and report case studies using specific architectures [23][18]. However, to make progress on big geospatial data analysis, we need to engage the large community of remote sensing researchers. In this paper, we consider how big EO data architectures can support the needs of data analytics. Our paper examines ways to cut the effort required for researchers to develop and validate algorithms for extracting information for big EO data.

We take the viewpoint that architectures should serve applications, and not the other way around. To clarify the researcher's problem, we consider the needs for an important

A METHOD TO ESTIMATE TEMPORAL INTERACTION IN A CONDITIONAL RANDOM FIELD BASED APPROACH FOR CROP RECOGNITION

P. M. A. Diaz [a], R. Q. Feitosa [a,c,*], I. D. Sanches [b], G. A. O. P. Costa [c]

[a] Pontifical Catholic University of Rio de Janeiro, Brazil – (pmad9589, raul)@ele.puc-rio.br
[b] National Institute for Space Research, Brazil – ieda.sanches@inpe.br
[c] Rio de Janeiro State University, Brazil – gilson.costa@ime.uerj.br

Commission VII, WG VII/4

KEY WORDS: Conditional Random Fields, Crop Recognition, Multitemporal Image Analysis.

ABSTRACT:

This paper presents a method to estimate the temporal interaction in a Conditional Random Field (CRF) based approach for crop recognition from multitemporal remote sensing image sequences. This approach models the phenology of different crop types as a CRF. Interaction potentials are assumed to depend only on the class labels of an image site at two consecutive epochs. In the proposed method, the estimation of temporal interaction parameters is considered as an optimization problem, whose goal is to find the transition matrix that maximizes the CRF performance, upon a set of labelled data. The objective functions underlying the optimization procedure can be formulated in terms of different accuracy metrics, such as overall and average class accuracy per crop or phenological stages. To validate the proposed approach, experiments were carried out upon a dataset consisting of 12 co-registered LANDSAT images of a region in southeast of Brazil. Pattern Search was used as the optimization algorithm. The experimental results demonstrated that the proposed method was able to substantially outperform estimates related to joint or conditional class transition probabilities, which rely on training samples.

1. INTRODUCTION

Remote sensing (RS) data has been increasingly applied to assess agricultural yield, production, and crop condition. Single date classification is inappropriate for this purpose, as the spectral appearance changes over time as crops evolve through their characteristic phenological circles.

Conditional Random Fields (CRF) have deserved considerable attention of the scientific community in the recent years for crop recognition from multitemporal images, mainly due to its ability to model interactions of neighbouring image sites both in the spatial and temporal domains. These two forms of interactions are quite different in nature and the strategies proposed so far to model them are similarly diverse. In the present work we concentrate on the temporal interactions alone.

Methods for multitemporal image analysis can be grouped into three main categories (Hoberg et al., 2011). The first one is related to the classification of single images based on a single powerful classifier or on a combination of classifiers. This approach does not take into account the temporal dependencies (Bruzone et al., 2004) (Waske and Braun, 2009). The second one is based on modelling temporal dependencies by rules (Simonneaux et al., 2008), or adaptive strategies to select the relevance of features over time for specific crops (Müller et al., 2010). The last one incorporates temporal dependencies into statistical models (Melgani and Serpico, 2004) (Leite et al., 2011).

Approaches that take the temporal dependencies into account usually model temporal interaction by class transition matrices that can be determined by an expert (Hoberg et al., 2010) (Hoberg et al., 2011) empirically from existing data sources, or computed statistically (Leite et al., 2011) (Kenduiywo et al., 2015).

In (Hoberg et al., 2015) and (Liu et al. 2008), temporal interactions are represented by transition matrices I_d whose elements are related to conditional probabilities, in other words, $I_d(i,j)$ is related to the probability of an image site belonging to class ω_j at epoch t_{d+1}, given that it belongs to class ω_i at epoch t_d, whereby ω_i and ω_j are class labels and t_d and t_{d+1} are adjacent epochs in the multitemporal data set.

In contrast, other works model interaction potentials as joint probabilities (e.g., Niemeyer et al., 2014). In the present context, $I_d(i,j)$ would be related in this case to the probability of an image site belonging to class ω_i in epoch t_d and to ω_j in epoch t_{d+1}.

Both approaches lead to theoretically plausible estimates of the interaction potential, but with no guarantee that it maximizes the classification accuracy.

In this work we address this issue. We propose a supervised method to estimate the temporal interaction in a CRF based framework for crop recognition. Starting from a class transition matrix computed upon training samples either as conditional or joint probabilities, the method fine tunes the estimate so as to maximize the CRF classification accuracy.

The remainder of the paper is organized as follows. Section 2 presents the CRF based approach used in the experiments as well as the proposed method. Section 3 presents and discusses the results obtained in the experiments. Finally, Section 4 summarizes the conclusions obtained in this work and indicate future directions.

* Corresponding author

IMPROVEMENT EVALUATION ON CERAMIC ROOF EXTRACTION USING WORLDVIEW-2 IMAGERY AND GEOGRAPHIC DATA MINING APPROACH

V. S. Brum-Bastos [a]*, B. M. G. Ribeiro [b], C. M. D. Pinho [c], T. S. Korting [d], L. M. G. Fonseca [d]

[a] University of St. Andrews, Department of Geography and Sustainable Development, St. Andrews – Fife, United Kingdom – vdsbb@st-andrews.ac.uk

[b] Federal University of Rio Grande do Sul (UFRGS), Graduate Program on Urban and Regional Planning (PROPUR) Porto Alegre – RS, Brazil – barbara.giaccom@ufrgs.br

[c] Federal University of ABC (UFABC), Engineering, Modeling and Applied Social Sciences Center (CECS), São Bernardo do Campo – SP, Brazil – carolina.pinho@ufabc.edu.br

[d] National Institute for Space Research (INPE), Image Processing Division (DPI), São José dos Campos – SP, Brazil – (tkorting, leila)@dpi.inpe.br

Commission ThS15

KEY WORDS: Geographical Data Mining, GEOBIA, WorldView-2, Ceramic roof, C4.5, Decision Tree, Classification accuracy.

ABSTRACT:

Advances in geotechnologies and in remote sensing have improved analysis of urban environments. The new sensors are increasingly suited to urban studies, due to the enhancement in spatial, spectral and radiometric resolutions. Urban environments present high heterogeneity, which cannot be tackled using pixel–based approaches on high resolution images. Geographic Object–Based Image Analysis (GEOBIA) has been consolidated as a methodology for urban land use and cover monitoring; however, classification of high resolution images is still troublesome. This study aims to assess the improvement on ceramic roof classification using WorldView-2 images due to the increase of 4 new bands besides the standard "Blue-Green-Red-Near Infrared" bands. Our methodology combines GEOBIA, C4.5 classification tree algorithm, Monte Carlo simulation and statistical tests for classification accuracy. Two samples groups were considered: 1) eight multispectral and panchromatic bands, and 2) four multispectral and panchromatic bands, representing previous high-resolution sensors. The C4.5 algorithm generates a decision tree that can be used for classification; smaller decision trees are closer to the semantic networks produced by experts on GEOBIA, while bigger trees, are not straightforward to implement manually, but are more accurate. The choice for a big or small tree relies on the user's skills to implement it. This study aims to determine for what kind of user the addition of the 4 new bands might be beneficial: 1) the common user (smaller trees) or 2) a more skilled user with coding and/or data mining abilities (bigger trees). In overall the classification was improved by the addition of the four new bands for both types of users.

1. INTRODUCTION

Novel development in remote sensing technologies have enhanced urban land use and land cover mapping over the last two decades, especially due to the availability of high–resolution images (Blaschke, 2010). The sensors aboard new satellites are increasingly suited to urban studies, due to the enhancement in spatial, spectral and radiometric resolutions (Pinho et al., 2012; Ribeiro et al., 2011). As a result, sub–metric objects have been discriminated, which widely benefits urban studies using remote sensing data. The information extracted from these products are of great importance on the development of medium and long–term investments planning, monitoring the increasing demand for infrastructure and social equipment, and supporting public policies in compliance with environmental guidelines and targeted to provide better quality of life to society.

Recent advances in geotechnologies provide resources to propose innovative strategies for urban and environmental management, including remote sensing data and computational resources for processing them, which, together, are able to generate high–quality databases and maps.

Complex urban environments present high heterogeneity, which cannot be tackled using pixel–based approaches on high resolution images. The solely use of spectral information is insufficient to describe different types of settlements due to variation in the structure, material and shape. Hence more refined image analysis methods are being successfully applied for urban studies using high spatial resolution data.

Geographic Object–Based Image Analysis (GEOBIA) has been consolidated as an efficient methodology for urban land use and land cover monitoring. However, even after continuous advances in GEOBIA, classification of high resolution images is still troublesome (Hay & Castilla, 2006). Softwares that perform GEOBIA provide a great number of attributes and different ways to model the semantic network, which make the task of classification lengthy and complex (Korting et al., 2008; Ribeiro & Fonseca, 2013). Determining the most relevant features to be used in classification routines is not always an easy task when conventional exploratory analyses are carried out (e.g., scatter plot, histograms, feature values shown in grey levels, etc.). Research concerning GEOBIA has presented innumerous advances, though some problems related to the feature selection and large amount of data are still unsolved.

* Corresponding author

This contribution has been peer-reviewed.
doi:10.5194/isprsarchives-XLI-B7-883-2016

USING LANDSAT 8 IMAGE TIME SERIES FOR CROP MAPPING IN A REGION OF CERRADO, BRAZIL

H. do N. Bendini [a], I. D. Sanches [a], T. S. Körting [b], L. M. G. Fonseca [b], A. J. B. Luiz [c], A. R. Formaggio [a]

[a] DSR/INPE, Remote Sensing Division of the National Institute for Space Research – (hbendini, formag)@dsr.inpe.br, ieda.sanches@inpe.br
[b] DPI/INPE, Image Processing Division of the National Institute for Space Research – tkorting@dpi.inpe.br, leila.fonseca@inpe.br
[c] Embrapa Environment, Brazilian Agricultural Research Corporation- alfredo.luiz@embrapa.br

Commission VI, WG VI/4

KEY WORDS: Crop mapping, Landsat Time series, Random Forest algorithm

ABSTRACT:

The objective of this research is to classify agricultural land use in a region of the Cerrado (Brazilian Savanna) biome using a time series of Enhanced Vegetation Index (EVI) from Landsat 8 OLI. Phenological metrics extracted from EVI time series, a Random Forest algorithm and data mining techniques are used in the process of classification. The area of study is a region in the Cerrado in a region of the municipality of Casa Branca, São Paulo state, Brazil. The results are encouraging and demonstrate the potential of phenological parameters obtained from time series of OLI vegetation indices for agricultural land use classification.

1. INTRODUCTION

1.1 Overview

Agriculture has significant participation in the Brazilian economy; it is the main responsible for the positive trade balance of the country. Given the high availability of arable land, and taking into account the growing demand for food in the world, Brazil has been consolidate as a big player in the world agricultural scenario. On the other hand, sustainable agriculture expansion is a key concern in Brazilian Cerrado biome, which suffers a great land conversion pressure. In this way, crop mapping is strategic to estimate acreage and production, as well as to better understanding the distribution of croplands, and its impact on the environment. In this context, remote sensing is an important tool due to its ability to generate information on large scale in a cost-effective way. With advances in data processing and storage technologies, and the availability of long-term image series, remote sensing is undergoing a paradigm shift, in which time series techniques allow to consider seasonal variations of the analysed target. This approach is useful for vegetation studies, especially in agricultural areas, since vegetation cover is quite dynamic in time, and the ability to capture these variations is essential to discriminate different types of crops, through its phenological characteristics. For this purpose, most studies have explored time series of vegetation indices like NDVI or EVI from MODIS data (Jakubauskas et al., 2002; Sakamoto et al., 2005; Wardlow et al., 2007; Esquerdo et al., 2011; Arvor et al., 2011; Körting, 2012; Risso et al., 2012; Coutinho et al., 2013; Borges & Sano, 2014; Tomás et al., 2015; Neves et al., 2016). However, there is a demand to produce more detailed maps, which can be obtained from higher spatial resolution satellites such as Landsat-like (Zheng et al., 2015; Peña et al., 2015; Pan et al., 2015). Therefore, the objective of this study is to employ phenological metrics obtained from time series of Landsat 8/OLI vegetation indices to classify agricultural land use in the municipality of Casa Branca, located in the Cerrado (Brazilian Savanna) biome in the state of São Paulo, Southeast region from Brazil.

1.2 Study area

We conducted our study in the south of Casa Branca municipality, in state of São Paulo, Brazil (Figure 1). Such region is located in an overlapping area of two adjacent Landsat Worldwide Reference System 2 (WRS – 2) scenes (path/row 219/75 and 220/75), providing a temporal resolution of 8 days (Luiz et al., 2015). Casa Branca city has a tropical wet and dry climate (Aw, according to the Köppen Climate Classification System) with average annual temperature of 21.5º C and a seasonal rainfall pattern with most rainfall occurring from October to March. The average annual precipitation is 109.18 mm (CEPAGRI/UNICAMP, 2016).

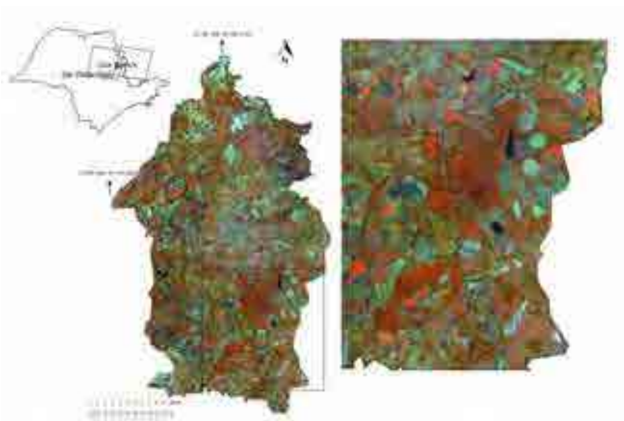

Figure 1. False color (bands 5, 6 and 4 in red, green and blue respectively) OLI Landsat imagery of the study area

In this region, farmers grow a variety of crops along all the year. Major field crops in this area are sugarcane, corn, bean, potato, soybean, peanuts, sorghum and cassava. There are also a

doi:10.5194/isprsarchives-XLI-B8-845-2016

Using dynamic geospatial ontologies to support information extraction from big Earth observation data sets

Gilberto Camara[1], Adeline Maciel[1], Victor Maus[1], Lubia Vinhas[1], and Alber Sanchez[1]

[1]Image Processing Division, National Institute for Space Research, Av dos Astronautas 1758, Sao Jose dos Campos, 12227-001 Brazil.
Email: {gilberto.camara, adeline.maciel, victor.maus, lubia.vinhas, alber.ipia}@inpe.br

Abstract

This paper presents the spatiotemporal interval logic formalism and shows how to use it for reasoning about land use change using big Earth observation data. This formalism improves our ability to extract information from large land remote sensing data sets.

1 Events as key concepts for describing land use change

Remote sensing satellites are the only source that provides consistent data about the Earth's land and oceans. The open availability of big Earth observation data has led to an opportunity to improve information on land changes in the planet. However, most studies that use remote sensing images to detect change still adopt a *snapshot* approach. Image from a sequence are classified one by one; results are compared to account for change. There is no actual representation of the occurrences of change, but only of their effects. Two land areas with different change trajectories whose initial and final states are the same cannot be distinguished. With access to big data sets, researchers need better ways to describe and understand change. The challenge is to make best use of big Earth observation data sets to represent change.

This paper uses the concept of *'events'* from dynamic spatial ontologies to describe land use change. Events are complete entities on their respective time intervals; their lifetime is limited while objects persist in time and are complete in space (Galton and Mizoguchi, 2009; Worboys, 2005; Hacker, 1982). Since events are intrinsically related to the objects they modify, a geospatial event calculus should specify not only what happens, but also which objects are affected by such changes. We present an event calculus formalism for reasoning about land use change. The formalism is general enough to be applied in other geospatial domains.

To define events in big Earth observation data sets, multiple satellite observations of an area are mapped to 3D arrays in space-time. A pixel location (x, y) in consecutive times $t_1, ..., t_m$ makes up a satellite image time series (Figure 1a). One can extract land use change information

HOW TO EFFECTIVELY OBTAIN METADATA FROM REMOTE SENSING BIG DATA?

Körting, T. S.,* Namikawa, L. M., Fonseca, L. M. G., Felgueiras, C. A.

National Institute for Space Research – INPE Brazil
DPI – Image Processing Division at the OBT – General Earth Observation Coordination
Av. dos Astronautas,1758 São José dos Campos, Brazil
{thales.korting, laercio.namikawa, leila.fonseca, carlos.felgueiras}@inpe.br

KEY WORDS: Big data, Remote sensing, Metadata, Image processing, Vegetation indices, Water indices, Pattern recognition

ABSTRACT:

What can be considered big data when dealing with remote sensing imagery? In general terms, big data is defined as data requiring high management capabilities characterized by 3 V's: Volume, Velocity and Variety. In the past, (e.g. 1975), considering the computational and databases resources available, a series of Landsat-1 imagery from the same region could be considered big data. Nowadays, several satellites are available, and they produce massive amounts of data. Certainly, an image data set obtained by a single satellite, for a specific region and along time, fills the 3 V's requirements to be considered big data as well. In order to deal with remote sensing big data, we propose to explore the generation of metadata based on the detection of simple features. Besides the intrinsic geographic information on every remote sensing scene, no additional metadata is usually considered. We propose basic image processing algorithms to detect basic well-known patterns, and include them as tags, such as cloud, shadow, stadium, vegetation, and water, according to what is detectable at each spatial resolution. In this work we show preliminary results using imagery from RapidEye sensor, with 5 meter spatial resolution, composed by two full coverages of Brazil with RapidEye multispectral imagery (around 40k scenes).

1. INTRODUCTION

Nowadays, several references to the term big data are available, many of them without a proper understanding of what is the real meaning of it. Since 2001, Big data has been defined as data requiring high management capabilities characterized by 3 V's: Volume, Velocity and Variety, as proposed by (Laney, 2001). In (Plunkett et al., 2013), the authors provide examples of what has been considered big data so far. Examples include Web server and application logs, digital video and music, clickstream data, social networks, smartphone location-based services, real-time trading data, blogs and social media. It is clear that most of the given examples are Internet-related, however we can find other examples of big data generation, such as remote sensing.

Remote sensing satellites fill the requirements to be characterized as big data. Since at each day new images are obtained, and previously captured images are also combined as time series, it is possible to confirm the constant growing volume and also the increasing velocity of data gathering. New satellites are being launched and their design life (duration) are frequently overcome. One great example is the Landsat 5 satellite, projected with a 5-year design life, that returned scientifically viable data for 28 years (USGS, 2016). In terms of variety, remote sensing data is expanding the amount of spectral channels (i.e. Landsat 5 and 7 produces images in 8 spectral channels; Landsat 8, in 11 channels), which means different ways to capture spectral interaction between targets and eletromagnetic radiation. When we focus the analysis in the GEOBIA approach, the variety of data related to the same target increases more, since with the use of spectrally homogeneous regions, we combine the intra-region spectral information, such as average pixel values or texture, with spatial information, such as geometric features, and also relations to the neighborhood. Considering the aforementioned reasons, it is clear that remote sensing is a source of big data.

In this paper, we propose a method to work in a set of images composed by two full coverages of Brazil with RapidEye multispectral imagery (MMA, 2016), from 2012 and 2014. RapidEye images are generated from a constellation of 5 satellites located at the same orbital plane, and carrying the same sensors (BlackBridge, 2015). Available RapidEye imagery are processed into level 3A, which corresponds to geometric, radiometric and sensor correction, and mosaicked into 25 by 25 km tiles with a 5 meter pixel size, created from the acquisition sampled at 6.5 meters at the nadir. The multispectral bands are 5: blue ($0.44-0.51\mu$m), green ($0.52-0.59\mu$m), red ($0.63-0.685\mu$m), red edge ($0.69-0.73\mu$m), and near infra-red ($0.76-0.85\mu$m).

2. METHODOLOGY

In this section we describe our proposal to deal with remote sensing big data for metadata generation, which is depicted in the diagram of Figure 1. We also provide references for the used indices and algorithms to detect the presence of target patterns in images.

Our proposal is to integrate a set of simple algorithms for pattern recognition (blocks called Detector for pattern 1 ... N in Figure 1) without a strong compromise with accuracy, therefore we can consider these algorithms as weak detectors. Our expected level of metadata to be generated is as superficial as tags like cloud, shadow, stadium, vegetation, and water, according to what is detectable in each spatial resolution.

The basic idea in dealing with remote sensing big data plus a stream of incoming images, is to provide a continuous workflow, which means a system that keeps running (see the block New images), allowing the insertion of algorithms for detecting new patterns on-the-fly (block New detectors). With the provided structure, the algorithm can run more than once for the same image, which is useful when parameters are changed or new algorithms are inserted.

2.1 Pre-processing

When dealing with remote sensing big data, which should include heterogeneous sources of images, specific parameters for each

*Corresponding author

FIRST EXPERIMENTS USING THE IMAGE FORESTING TRANSFORM (IFT) ALGORITHM FOR SEGMENTATION OF REMOTE SENSING IMAGERY

A. R. Soares[a], [*]T. S. Körting[a], L. M. G. Fonseca[a]

[a]Image Processing Division, National Institute of Space Research, INPE, São José dos Campos, Brazil
(anderson.soares,thales.korting,leila.fonseca)@inpe.br

KEY WORDS: Image Segmentation, Image Foresting Transform, Multiresolution Segmentation, Segmentation Comparison

ABSTRACT:

Image segmentation is a traditional method in Remote Sensing and a fundamental problem in image processing applications. It has been widely used, especially with the emergence of the Geographic Object-Based Image Analysis (GEOBIA). The results of segmentation must create uniform areas, which must allow a simpler interpretation by the users and simpler representation for classification algorithms. Several algorithms were proposed through the years, using different approaches. One that is widely used in Remote Sensing applications is the Multiresolution algorithm, that is based on the region growing method. Other, which has great potential and is applied in other research areas, is available on the Image Foresting Transform (IFT) framework, which has several image operators developed primarily for medical images. The Watershed from Grayscale Marker operator uses an edge image to perform the segmentation, however, we propose an extension of the edge detection algorithm, by summing normalized gradients of each band. This work aims to evaluate and compare these two segmentation algorithms, by comparing their results through supervised segmentation from reference regions, that were defined manually by an expert user. Quality measures were evaluated by four metrics, that represent the positional adjustment based the center of gravity, intensities, size, and the amount of overlap between the segment created by the algorithms and the reference segment. We selected 21 objects of a WorldView-2 multispectral image that were used to compute the metrics. Both methods reached similar results, by comparing the aforementioned 4 metrics applied to the 21 reference regions, IFT achieved better results for majority of regions. The IFT generated segments with similar shape when compared with the references, and the multiresolution generated results with similar sizes and positional adjustments. It may be explained by the fact that IFT uses an edge image to perform the segmentation. Both algorithms obtained similar agreement for intensity.

1. INTRODUCTION

Image segmentation is a fundamental problem in image processing, (Soille, 1999) defined segmentation as a process to split an image grouping the pixels by a similar attribute, such as the grey level, so the line which splits the areas, ideally, must be an edge. According to (Körting, 2012) the segmentation is one of the most challenging tasks in digital image processing.

Geographic Object-Based Image Analysis (GEOBIA) is devoted to developing automated methods to partition Remote Sensing imagery into meaningful image-objects, and assessing their characteristics through spatial, spectral and temporal scales (Hay and Castilla, 2008). It became widely used because it offers the potential to exploit geographical information.

The development of GEOBIA required a search for new methods for image segmentation. During the last decades, numerous techniques have been developed and applied in Remote Sensing analysis (Bins et al., 1996, Hay et al., 2005). The image segmentation algorithm that generates appropriate results has been the Multiresolution Segmentation (MRS) (Baatz and Schäpe, 2000), which probably is the most popular algorithm applied in GEOBIA.

Besides MRS good results, the search for new algorithms for image segmentation is still a necessity, and can not be limited by the MRS popularity. Different approaches may, also, produce similar results or even better results.

The Image Foresting Transform is a general tool for the design, implementation, and evaluation of image processing operators based on connectivity (Falcão et al., 2004), it defines a minimum-cost path forest in a graph, whose nodes are the image pixels and whose arcs are defined by an adjacency relation between pixels. The cost of a path is determined by a path-cost function f, which usually depends on local image properties (color, gradient, or pixel position) along the path between the nodes and the root.

*Corresponding author

The IFT Watershed from Grayscale Marker (called Watergray from this point), is one IFT operator in which the segmentation is computed from a gray scale image. According (Lotufo et al., 2002) it puts together several steps used in classical watershed morphological segmentation strategies in a single algorithm.

This paper evaluates and compares the Watergray operator, adapted to the Remote Sensing context, with the MRS algorithm.

2. METHODOLOGY

We compared the algorithms using 4 quality measures applied in 21 selected regions (Figure 1), on a WorldView-2 Multispectral image.

As the Watergray uses an edge image to perform the segmentation we implemented on MATLAB an adaptation of the traditional way to compute the gradient of the image. As described in Equation 1 we computed the traditional gradient for each band and normalized each one. With this preprocessing we expect that each band has the same importance when all gradients are summed.

$$\bigtriangledown_{image} = \sum_{i}^{N} w_N * \bigtriangledown_N \quad (1)$$

DETECTING ATLANTIC FOREST PATCHES APPLYING GEOBIA AND DATA MINING TECHNIQUES

C. D. Girolamo Neto [a], A. C. M. Pessôa [b], T. S. Körting [a], L. M. G. Fonseca [a]

[a] Image Processing Division – National Institute for Space Research – INPE – cesare@dsr.inpe.br; (thales; leila)@dpi.inpe.br
[b] Remote Sensing Division – National Institute for Space Research – INPE – ana.pessoa@inpe.br
Av. dos Astronautas, 1758, São José dos Campos, SP, Brazil.

KEY WORDS: Land cover, Classification, Landsat-8, Random Forest, Artificial Neural Networks, Feature selection

ABSTRACT:

Brazilian Atlantic Forest is one of the most devastated tropical forests in the world. Considering that approximately only 12% of its original extent still exists, studies in this area are highly relevant. In this context, this study maps the land cover of Atlantic Forest within the Protected Area of 'Macaé de Cima', in Rio de Janeiro State, Brazil, combining GEOBIA and data mining techniques on an OLI/Landsat-8 image. The methodology proposed in this work includes the following steps: (a) image pan-sharpening; (b) image segmentation; (c) feature selection; (d) classification and (e) model evaluation. A total of 15 features, including spectral information, vegetation indices and principal components were used to distinguish five patterns, including *Water*, *Natural forest*, *Urban area*, *Bare soil/pasture* and *Rocky mountains*. Features were selected considering well-known algorithms, such as Wrapper, the Correlation Feature Selection and GainRatio. Following, Artificial Neural Networks, Decision Trees and Random Forests classification algorithms were applied to the dataset. The best results were achieved by Artificial Neural Networks, when features were selected through the Wrapper algorithm. The global classification accuracy obtained was of 98.3%. All the algorithms presented great recall and precision values for the Natural forest, however the patterns of Urban area and Bare soil/pastures presented higher confusion.

1. INTRODUCTION

Land use and cover (LUC) analysis can be used to determinate how a specific area is being used, highlighting the anthropogenic interactions with the environment. In order to access patterns of LUC changes, it is vital to use data from remote sensing imagery (Brannstrom *et al.*, 2008). This technology allows the generation of LUC maps, showing areas being occupied by pastures, crops, natural vegetation, river courses and other features. They can also indicate areas of risk or those heavily degraded.

One of the most devastated Brazilian biomes is the Atlantic Forest. The second largest Brazilian forest has only 12% of its initial extent preserved (Ribeiro *et al.*, 2009). A large part of the occupation of this biome has occurred due to the expansion of urban centers and agricultural areas. Among Rio de Janeiro and São Paulo states, most of Atlantic Forest patches are usually in Protected Areas (PA's). LUC analysis on these territories is even more important when considering the possibility of degrading preserved natural vegetation areas (Figueroa & Sánchez-Cordero, 2008).

A procedure used to perform the LUC classification is the Geographic Object-Based Image Analysis (GEOBIA), which aims to classify an image based on similar characteristics of its objects. In addition to the spectral properties, GEOBIA can evaluate features associated with the shape, texture, contextual and semantic relationships of objects, increasing the chances of a more reliable classification (Camargo *et al.*, 2009).

Another methodology that has been constantly used on image classification is the Data Mining (DM). DM helps GEOBIA in the process of identifying patterns on objects. In this context, some DM classification algorithms have been used on LUC analysis, for instance: Decision Trees were used for vegetation mapping (Colstoun *et al.*, 2003), temporal analysis of agricultural crops (Körting, 2012) and urban LUC (Pinho *et al.*, 2012). Random Forests were applied to classify LUC on various locations (Smith, 2010; Müller *et al.*, 2015). Artificial Neural Networks were used assessing Natural vegetation LUC (Moreira *et al.*, 2013) and Agricultural areas (Andrade *et al.*, 2013). The choice to use each algorithm requires the analysis of the problem. The results of Decision Trees are easy to visualize (Witten *et al.*, 2011), Random Forests can avoid overfitting, and are also not very sensitive to noisy data (Breiman, 2001). The main advantage of Artificial Neural Networks is to solve complex problems (Haykin, 2009) and may outperform other classifiers on LUC classification (Song *et al.*, 2012).

Considering the importance of LUC information on Atlantic Forest areas and the potential of GEOBIA and DM techniques on image classification, this study aims to map the land use and cover of Atlantic Forest within the PA of 'Macaé de Cima', in Rio de Janeiro State, Brazil, using GEOBIA and DM techniques on an OLI/Landsat-8 image.

2. METHODOLOGY

2.1 Study site and Data

The study site is the PA of Macaé de Cima – Rio de Janeiro State (Figure 1). It is located between the coordinates of 22°17´S-22°27´S and 42°35´W-42°12´W, on the municipalities of Macaé and Nova Friburgo. It has an area of 350.000 square meters on steep region with rocky mountains and small valleys, with 72% of Atlantic Forest cover (INEA, 2007).

Data was obtained from the Operational Line Imager (OLI) sensor from Landsat-8 satellite, path/row 216/76 acquired on 10/14/2014.

Big data streaming for remote sensing time series analytics using MapReduce

Luiz Fernando Assis[1], Gilberto Ribeiro[1], Karine Reis Ferreira[1], Lúbia Vinhas[1], Eduardo Llapa[1], Alber Sanchez[1], Victor Maus[1], Gilberto Câmara[1]

[1]Image Processing Department
INPE - National Institute for Space Research
Av dos Astronautas 1758
Caixa Postal Sao Jose dos Campos – SP – Brazil

{luizffga,gribeiro,karine,lubia,edullapa,alber.ipia,gilberto}@dpi.inpe.br

Abstract. *Governmental agencies provide a large and open set of satellite imagery which can be used to track changes in geographic features over time. The current available analysis methods are complex and they are very demanding in terms of computing capabilities. Hence, scientist cannot reproduce analytic results because of lack of computing infrastructure. Therefore, we propose a combination of streaming and map-reduce for time series analysis of time series data. We tested our proposal by applying the classification algorithm BFAST to MODIS imagery. Then, we evaluated account computing performance and requirements quality attributes. Our results revealed that the combination between Hadoop and* ***R*** *can handle complex analysis of remote sensing time series.*

1. Introduction

Currently, there is huge amount of remote sensing images openly available, since many space agencies have adopted open access policies to their repositories. This large data sets are a good chance to broaden the scope of scientific research that uses Earth observation (EO) data. To support this research, scientists need platforms where they can run algorithms that analysis big Earth observation data sets. Since most scientists are not data experts, they need data management solutions that are flexible and adaptable.

To work with big EO, we need to develop and deploy innovative knowledge platforms. When users want to work with hundreds or thousands of images to do their analysis, it is not practical to work with individual files at their local disks. Innovative platforms should allow scientists to perform data analysis directly on big data servers. Scientists will be then able to develop completely new algorithms that can seamlessly span partitions in space, time, and spectral dimensions. Thus, we share the vision for big scientific data computing expressed by the late database researcher Jim Gray: *"Petascale data sets require a new work style. Today the typical scientist copies files to a local server and operates on the data sets using his own resources. Increasingly, the data sets are so large, and the application programs are so complex, that it is much more economical to move the end-user's programs to the data and only communicate questions and answers rather than moving the source data and its applications to the user's local system"* [Gray et al. 2005].

For instance, the standard for land use and land cover monitoring includes to select and download a set of images, processing of each one using visual interpretation or

Assessment of a Multi-Sensor Approach for Noise Removal on Landsat-8 OLI Time Series Using CBERS-4 MUX Data to Improve Crop Classification Based on Phenological Features

Hugo N. Bendini[1], Leila M. G. Fonseca[2], Thales S. Körting[2], Rennan F. B. Marujo[2], Ieda D. Sanches[1], Jeferson S. Arcanjo[2]

[1]Divisão de Sensoriamento Remoto – Instituto Nacional de Pesquisas Espaciais (INPE)
[2]Divisão de Processamento de Imagens – INPE
Caixa Postal 515 – 12.227-010 – São José dos Campos – SP – Brazil

hbendini@dsr.inpe.br, {leila.fonseca,thales.korting, ieda.sanches}@inpe.br, jeferson@dpi.inpe.br

Abstract. *We investigated a method for noise removal on Landsat-8 OLI time-series using CBERS-4 MUX data to improve crop classification. An algorithm was built to look to the nearest MUX image for each Landsat image, based on user defined time span. The algorithm checks for cloud contaminated pixels on the Landsat time series using Fmask and replaces them with CBERS-4 MUX to build the integrated time series (Landsat-8 OLI+CBERS-4 MUX). Phenological features were extracted from the time series samples for each method (EVI and NDVI original time series and multi-sensor time series, with and without filtering) and subjected to data mining using Random Forest classification. In general, we observed a slight increase in the classification accuracy when using the proposed method. The best result was observed with the EVI integrated filtered time series (78%), followed by the filtered Landsat EVI time series (76%).*

1. Introduction

Given the large availability of arable land, and the growing demand for food in the world, Brazil has been consolidated as a big player on the global agricultural scene. Remote sensing is an important tool used within agriculture, regarding its ability to generate information on a large scale in a cost-effective way. In this way, agricultural mapping has become strategic enabling to provide better understanding of the distribution of croplands, and its impact on the environment. With advances in data processing and storage technologies as well as the availability of consistent and continuous long-term image series, remote sensing is undergoing a paradigm shift. Time series techniques stand out for allowing seasonal variation accounts of the analyzed target. Although the use of time series for cropland classification has been well explored using MODIS data (Sakamoto et al., 2005; Arvor et al., 2011; Körting, 2012; Risso et al., 2012; Borges & Sano, 2014; Neves et al., 2016), there is still a demand for more detailed maps, which are made possible from time series with finer spatial resolutions, such as Landsat-like images (Zheng et al., 2015; Peña et al., 2015; Pan et al., 2015; Bendini et al., 2016). As the temporal resolution of Landsat-like satellites it is still low (16 days, generally), an open question in the scientific literature is about how to deal

Assessment of texture features for Brazilian savanna classification: a case study in Brasilia National Park

Cesare Di Girolamo Neto[1], Leila Maria Garcia Fonseca[1], Thales Sehn Körting[1]

[1]Instituto Nacional de Pesquisas Espaciais – INPE
Caixa Postal 515 - 12227-010, São José dos Campos – SP – Brazil

{cesare.neto;leila.fonseca; thales.korting}@inpe.br

Abstract. *Distinguishing Brazilian savanna physiognomies is an essential task to better evaluate carbon storage and potential emissions of greenhouse gases. In this study, we propose to evaluate the potential of texture features to improve the discrimination among five physiognomies in the Brazilian savanna: Open Grasslands, Shrubby Grassland, Shrubby Savanna, Savanna Woodland and Gallery Forest. Texture features extracted from RapidEye images and also from Spectral Linear Mixture Model components and Vegetation Index are evaluated in this study. Results showed that texture features based on GLCM can reduce misclassification for Open Grasslands, Shrubby Grasslands and Shrubby Savanna classes.*

1. Introduction

Brazilian savanna, also known as Cerrado, occupies an area of approximately two million square Kilometers on the Brazilian territory, mainly in the central part of Brazil (MMA, 2015). Cerrado is one of the richest biomes in the world and it contains more than 160.000 species of plants, animals and fungi (Ferreira *et al*., 2003). Besides that, Cerrado is responsible for storing about 5.9 billion tons of carbon in vegetation and 23.8 billion tons in the ground (MMA, 2014).

The loss of natural vegetation in Cerrado reached 45.5% of its original area by 2013 (MMA, 2015). The loss of biodiversity can lead to problems such as: soil erosion, water pollution, carbon cycle of instability, microclimate changes and also biome fragmentation (Klink & Machado, 2005). Considering these negative effects on biodiversity, it is essential to promote strategies to monitor the Cerrado biome.

Mapping of heterogeneous tropical areas, such as Cerrado, should be carried out considering biological, climatic and topographical information. The major natural formations in Cerrado are Grasslands, Shrublands and Woodlands (Figure 1). Their mapping has been the subject of several studies. Sano *et al*. (2009) performed visual interpretation of satellite images to produce maps of Cerrado. This process was very time consuming and difficult to discriminate Grasslands.

The difficulty to map Cerrado patterns is even greater when considering more formations than those mentioned above. For example, the system proposed by Ribeiro & Walter (2008) splits these major formations into 14 physiognomies. Identifying these physiognomies is important to evaluate carbon storage and potential emissions of greenhouse gases for each type of land cover.

Gerenciamento de nuvem de pontos em SGBD: avaliando a extensão PointCloud para PostgreSQL

Vitor C. F. Gomes[1], **Luciane Y. Sato**[2],
Gilberto Ribeiro de Queiroz[2], **Lúbia Vinhas**[2], **Karine Reis Ferreira**[2]

[1] Instituto de Estudos Avançados – Departamento de Ciência e Tecnologia Aeroespacial
CEP 12.228-001 – São José dos Campos – SP – Brasil

[2] Coordenação de Observação da Terra – Instituto Nacional de Pesquisas Espaciais
Caixa Postal 515 – CEP 12.227-010 – São José dos Campos – SP – Brasil

vitor@ieav.cta.br, lusato@dsr.inpe.br
{gilberto.queiroz, lubia.vinhas, karine.ferreira}@inpe.br

Abstract. *The potential observed in various applications and the advancement of technologies for point cloud data acquisition has increased the availability of this type of geospatial data. With increasing volume, new challenges arise in the efficient management of this information. Recently, the Database Management System PostgreSQL has supported this data type through the extension PointCloud. This paper explores this extension, through the analysis of its capabilities and its use in an experimental set of LiDAR data. Performance tests and disk usage metrics are collected for different compression methods. In addition, we present performance testing and disk usage metrics collected for different compression methods.*

Resumo. *O potencial observado em diversas aplicações e o avanço das tecnologias de aquisição de nuvem de pontos têm aumentado a disponibilidade desse tipo de dado geoespacial. Com o aumento do volume, novos desafios surgem quanto ao gerenciamento eficiente dessas informações. Recentemente, o Sistema Gerenciador de Banco de Dados PostgreSQL tem gerenciado esse tipo de dado através da extensão PointCloud. Este trabalho explora essa extensão, através da análise de suas capacidades e do seu uso em um conjunto experimental de dados LiDAR. Além disso, apresentamos testes de desempenho e métricas de utilização de disco coletadas para diferentes métodos de compressão.*

1. Introdução

Nuvens de pontos 3D representam uma categoria essencial de dados geoespaciais usados em uma variedade de aplicações e sistemas de geoinformação. O uso desse tipo de dado tem crescido ao longo da última década, com aplicações em diversas áreas do conhecimento, como em modelagens de objetos e construções, sítios arqueológicos, mapeamentos topográficos, aplicações florestais entre outros [Richter et al. 2015, Martinez-Rubi et al. 2016].

Tecnologias modernas de aquisição e processamento de nuvens de pontos 3D, como ecobatímetros multi-feixe, sistemas de acompanhamento de dados sísmicos e LiDAR (*Light Detection And Ranging*) em plataformas fixas ou móveis ou aeroembarcados, podem produzir de milhares a trilhões de pontos. Esses pontos, além dos dados de

Spatiotemporal Data Representation in R

Lorena A. Santos[1], Karine R. Ferreira[1], Gilberto R. Queiroz [1], Lubia Vinhas[1]

[1]National Institute for Space Research
Av. dos Astronautas, 1758,
12227-010 - São José dos Campos (SP) - Brazil

{lorena.santos,karine.ferreira,gilberto.queiroz,lubia.vinhas}@inpe.br

Abstract. *Recent advances in devices that collect geospatial information have produced massive spatiotemporal data sets. Earth observation and GPS satellites, sensor networks and mobile gadgets are examples of technologies that have created large data sets with better spatial and temporal resolution than ever. This scenario brings a challenge for Geoinformatics: we need software tools to represent, process and analyze these large data sets efficiently. R is a environment widely used for data analysis. In this work, we present a study of spatiotemporal data representation in R. We evaluate R packages to access and create three spatiotemporal data types as different views on the same observation set: time series, trajectories and coverage.*

1. Introduction

Recently, the amount of devices that collect geospatial information has greatly increased. Earth observation and GPS satellites, sensor networks and mobile gadgets are examples of technologies that have created large data sets with better spatial and temporal resolution than ever. This technological advance brings many challenges for Geoinformatics. We need novel software tools to represent, process and analyze big spatiotemporal data sets efficiently.

In Geoinformatics, spatiotemporal data representation is an open issue. Spatial information is represented following well-established models and concepts. This includes the dichotomy between object-based and field-based models [Galton 2004]. Examples of long-standing concepts are vector and raster data structures, topological operators, spatial indexing, and spatial joins [RIGAUX et al. 2002]. Most existing GIS and spatial database systems, such as PostGIS and Oracle Spatial, are grounded on these concepts. However, there is no consensus on how to represent spatiotemporal information in computational systems.

Many existing proposals of spatiotemporal data models focus on representing the evolution of objects and fields over time. Some proposals are specific for discrete changes in objects [Worboys 1994] [Hornsby and Egenhofer 2000], others for moving objects [Guting and Schneider 2005] [ISO 2008] and still others for fields or coverage [Liu et al. 2008] [OGC 2006]. To properly capture changes in the world, representing evolution of objects and fields over time is not enough. We also need to represent events and relationships between events and objects explicitly [Worboys 2005]. Events are occurrents [Galton and Mizoguchi 2009]. They are individual happenings with definite beginnings and ends. The demand for models that describe events has encouraged recent research on spatiotemporal data modeling [Galton and Mizoguchi 2009].

Bayesian network model to predict areas for sugarcane expansion in Brazilian Cerrado

Alexsandro C. de O. Silva, Leila M. G. Fonseca, Thales S. Körting

National Institute for Space Research (INPE)
Post Office Box: 515 – 12227-010 – São José dos Campos – SP – Brazil

{alexsandro.silva, leila.fonseca, thales.korting}@inpe.br

Abstract*: The growing demand for ethanol has powered the shift of the sugarcane frontier into the Brazilian Cerrado, mainly in the states of Goiás and Mato Grosso do Sul. Therefore, this study aims to propose Bayesian Network models for identifying potential areas for sugarcane expansion in Goiás and Mato Grosso do Sul states. The models take into account constraint factors in relation to the sugarcane expansion such as topography, soil aptitude, climate conditions, and available infrastructures. Results showed that Bayesian Network models proposed in this study were able to represent the tendency of sugarcane expansion.*

1. Introduction

The intense demand for ethanol in the last years has stimulated the sugarcane expansion in new areas, which intensify the competition with other agricultural areas, increase the land prices and reduce the options for further expansions [Adami et al. 2012; Granco et al. 2015]. The saturation of traditional producing areas such as the state of São Paulo [Castro et al. 2010; Shikida 2013] along with the higher production costs for the industry has motivated the producers to search potential new areas for production of sugarcane.

The ethanol industry identified in the Brazilian Cerrado opportunities for investment, especially in the states of Goiás and Mato Grosso do Sul – the Cerrado Biome covers 97% of Goias and 61% of Mato Grosso do Sul [BRASIL 2009]. However, even though favorable climate and soil conditions to the sugarcane cultivation [Shikida 2013] and affordable land prices, few mills were operating in the region. Until 2005, only 22 mills had been established in Goiás and Mato Grosso do Sul, which was a limiting factor for farmers to start planting sugarcane, once the crop needs to be promptly processed after harvesting [Granco et al. 2015].

To enhance the attractiveness for the sugarcane industry the government provided support through fiscal incentives, credit lines and investments in transportation infrastructure. As the result, the number of mills and areas planted to sugarcane in Goias and Mato Grosso do Sul increased approximately three times from 2005 to 2015 [Granco et al. 2015]. Consequently, theses states, which previously had an economy centered on cattle ranching and grains (soybean and corn), witnessed a strong sugarcane expansion [Shikida 2013].

The rapid sugarcane expansion and the eventual land cover changes in the Cerrado led the Brazilian government to implement the Sugarcane Agroecological Zoning to regulate the expansion and sustainable sugarcane production in Brazil, in 2009 [Manzatto et al. 2009]. To identify potential areas for sugarcane crop in the

PostGIS-T: towards a spatiotemporal *PostgreSQL* database extension

Rolf E. O. Simoes[1], **Gilberto Ribeiro de Queiroz**[1], **Karine Reis Ferreira**[1], **Lubia Vinhas**[1], **Gilberto Camara**[1]

[1]Instituto Nacional de Pesquisas Espaciais (INPE)
Caixa Postal 15.064 – 91.501-970 – São José dos Campos – SP – Brazil

{rolf.simoes, gilberto.queiroz, karine.ferreira,

lubia.vinhas}@inpe.br

Abstract. *The temporal dimension of spatial data has been the subject of discussion in the literature for a long time. While there are numerous Database Management System (DBMS) solutions only for spatial dimension, we did not observe the same situation for* spatiotemporal *data. Considering this gap, our purpose is to design and implement an extension to the DBMS PostgreSQL that is based on a formal spatiotemporal algebra in order to incorporate representations of spatiotemporal data within the DBMS. The proposed extension can be used in a large range of applications. We intend that this extension be a reasonable framework to store and handling observational remote sensing data usually present in applications like animal migration researches, wildfires monitoring, vessel tracking for monitoring fishing, and the like. In this work, we show how to apply it in a case study based on spatiotemporal data collected from drifting buoys belonging to the NOOA's Global Drifter Program.*

1. Introduction

Earth Observation data generation has been increased since the last decades. This phenomenon occurs, considering that a great amount of data are daily collected by different missions such as *CBERS*[1] in Brazil/China, *Landsat*[2] in the USA and *Sentinel*[3] in Europe. The development of mobile positioning technologies and its low costs are also factors that enable spatial data gathering through time.

These different data sources are associated with temporal dimension, mainly by allowing the monitoring of spatially located objects in time, either by allowing the time analysis by increasing the temporal resolution of the observations. The collection, representation and processing of this data have been largely facilitated by database management systems (DBMS) and their spatial extensions, which are based on international standards such as the OGC Simple Feature Specification [Herring 2011] and ISO geographic information standards [Kresse and Fadaie 2004]. Furthermore, while there are numerous DBMS solutions supporting the spatial dimension, we do not observe the same situation for spatiotemporal data.

[1]http://www.cbers.inpe.br/
[2]http://landsat.usgs.gov/
[3]https://sentinel.esa.int/

Web Services for Big Earth Observation Data

Lubia Vinhas[1], **Gilberto Ribeiro de Queiroz**[1],
Karine Reis Ferreira[1], **Gilberto Camara**[1]

[1]Instituto Nacional de Pesquisas Espaciais (INPE)
Av dos Astronautas, 1758, 12227-010, São José dos Campos, SP, Brazil

{lubia.vinhas, gilberto.queiroz}@inpe.br
{karine.ferreira, gilberto.camara}@inpe.br

Abstract. *The aim of geospatial web services is to enable users to share and process geospatial data on the Web, no matter what platform or protocol is used. In this paper, we investigate what are the design decisions needed to implement web services for big Earth observation (EO) data. The focus of the work is discussing what is unique about big EO data and why current standards are unsuitable for big EO data analytics. Instead, simpler APIs can be more effective for accessing big EO data than generic services such as WMS and WCS, specially for data analytics. We support this viewpoint by describing the WTSS Web Time Series Service that offers time series of remote sensing data using a simple API and is suited for big data analytics.*

1. Introduction

Earth observation (EO) satellites produce vast amounts of geospatial data. As an example, the archive of medium-resolution Landsat satellite at the US Geological Survey holds more than five million images of the Earth's surface, with about 1 PB of data. Most space agencies have adopted an open data policy, making unprecedented amounts of satellite data available for research and operational use. This data deluge has brought about a major challenge for Geoinformatics research: *How to design and build technologies that allow the EO community to use big data sets?*

EO satellites provide a continuous and consistent set of information about the Earth's land and oceans. Using big EO data sets, we can detect long-term changes and trends in the environment, and measure the impacts of climate change, urban and ocean pollution and land expansion for food production. To transform data in information, we need to analyse terabytes of data, with different spectral, temporal, and spatial resolutions, acquired by different sensors and stored at different places. Given the broad application areas of EO data, researchers need to consider how to store and organise them in a way that will facilitate data interoperability and data analytics.

The issue of handing big geospatial data has attracted much attention in recent literature. Vitolo et al. (2015) reviewed the web technologies dealing with big environmental data, concluding that domain-specific projects often require tailored solutions. Li et al. (2016) revisited the existing geospatial data handling methods and theories to determine if they are still capable of handling emerging geospatial big data. They concluded that traditional data handling approaches and methods are inadequate and new developments are needed. However, their conclusions are focused on the processing of discrete geographical data, as data captured by mobile devices and GPS based sensors. Amirian et al. (2014)

POSITION PAPER

FROM VGI TO CGI: COLLABORATIVE GEOGRAPHICAL INITIATIVES AS A BASIS FOR IMPROVED SPATIAL INFORMATION PRODUCTION

Gilberto Câmara (INPE)

The concept of "*voluntary geographical information*" was motivated by the rapid growth and widespread impact of Open Street Map and inspired by collective initiatives such as Wikipedia. The success of Open Street Map led to much expectation about the use of voluntary initiatives in other domains. Ideas such as "*Citizens as sensors*" and "*Citizen Science*" became widespread, leading to the expectation of a revolution on the production of geographical information. After a decade of their first inception, such predictions failed to become reality. There is no equivalent success history to Open Street Map. Even well-funded and supported VGI initiatives as the IIASA "Geo-Wiki" arguably have not delivered on its goals.

One possible explanation is that the VGI proponents failed to fully understanding the role of incentives. Projects such as Open Street Map, Wikipedia and Linux deliver tangible results: better maps, more knowledge, and improved software. Those involved know what they get in return. By contrast, initiatives such as "citizens as sensors" and "geo-wiki", in most cases, do not provide directly measurable benefits to volunteers. Part of the reason is the abstract nature of the such information. So, we need to conceive a new collaborative paradigm for geospatial data production.

We believe it is time for the VGI paradigm to be replaced, as far as more abstract information is produced, by the idea of CGI: *collaborative geographical initiatives*. The idea of CGI is to provide direct academic incentives to students and researchers who are involved in improving geographical information. The incentive is one that is well-established: authorship and citations in scientific papers.

Shared production of knowledge is becoming more and more widespread in sciences such as Physics and Genetics. The gravitational wave paper has more than 1,000 authors. In 2011, there were more than 400 indexed papers with more than 200 authors and more than 1000 papers with 50 authors or more. GIScience needs to draw inspiration from these disciplines.

Aluno (a): *Vivian Fróes Renó*

"VÁRZEAS AMAZÔNICAS: ALTERAÇÕES DA PAISAGEM E SEUS IMPACTOS NA PROVISÃO DE SERVIÇOS ECOSSISTÊMICOS E BEM-ESTAR DE COMUNIDADES RIBEIRINHAS"

Aprovado (a) pela Banca Examinadora em cumprimento ao requisito exigido para obtenção do Título de *Doutor(a)* em

Sensoriamento Remoto

Dra. Evlyn Márcia Leão de Moraes Novo

Presidente / Orientador(a) / INPE / SJCampos - SP

Dra. Maria Isabel Sobral Escada

Orientador(a) / INPE / SJCampos - SP

Dr. Camilo Daleles Rennó

Membro da Banca / INPE / SJCampos - SP

Dr. Yosio Edemir Shimabukuro

Membro da Banca / INPE / São José dos Campos - SP

Dra. Patrícia Fernanda do Pinho Koberle

Membro da Banca / USP / São Paulo - SP

Dra. Marie Paule Bonnet

Convidado(a) / UNB / Brasília - DF

Dr. Conrado de Moraes Rudorff

Convidado(a) / CEMADEN / Cachoeira Paulista - SP

Este trabalho foi aprovado por:

( ) maioria simples

(X) unanimidade

São José dos Campos, 19 de Agosto de 2016

Aluno (a): ***Vagner Luis Camilotti***

Título: "RECURSOS FLORESTAIS EXTRATIVISTAS EM COMUNIDADES NO SUDOESTE DO PARÁ: USO, IMPORTÂNCIA E CARACTERÍSTICAS DA PAISAGEM".

Aprovado (a) pela Banca Examinadora em cumprimento ao requisito exigido para obtenção do Título de ***Doutor(a)*** em

Ciência do Sistema Terrestre

Dra. Silvana Amaral Kampel

Presidente / INPE / SJCampos - SP

Dra. Maria Isabel Sobral Escada

Orientador(a) / INPE / SJCampos - SP

Dra. Patrícia Fernanda do Pinho Koberle

Orientador(a) / USP / São Paulo - SP

Dr. Antonio Miguel Vieira Monteiro

Membro da Banca / INPE / SJCampos - SP

Dra. Carolina Moutinho Duque de Pinho

Convidado(a) / Univ. Federal do ABC / São Paulo - SP

Dr. Nathan David Vogt

Convidado(a) / UNIVAP / São José dos Campos - SP

Este trabalho foi aprovado por:

( ) maioria simples

(X) unanimidade

São José dos Campos, 24 de Maio de 2016

Aluno (a): ***Victor Wegner Maus***

Título: "LAND USE AND ALND COVER MONITORING USING REMOTE SENSING IMAGE TIME SERIES".

Aprovado (a) pela Banca Examinadora em cumprimento ao requisito exigido para obtenção do Título de ***Doutor(a)*** em

Ciência do Sistema Terrestre

Dra. Silvana Amaral Kampel

Presidente / INPE / SJCampos - SP

Dr. Gilberto Câmara

Orientador(a) / INPE / SJCampos - SP

Dr. Fernando Manoel Ramos

Orientador(a) / INPE / São José dos Campos - SP

Dr. Dalton de Morisson Valeriano

Membro da Banca / INPE / SJCampos - SP

Dra. Maria Isabel Sobral Escada

Membro da Banca / INPE / SJCampos - SP

Dr. Daniel de Castro Victoria

Convidado(a) / EMBRAPA / Campinas - SP

Dr. Márcio Pupin de Mello

Convidado(a) / Boeing Res. & Tecno / São José dos Campos - SP

Este trabalho foi aprovado por:

( ) maioria simples

(x) unanimidade

São José dos Campos, 29 de Abril de 2016

Aluno (a): ***Anielli Rosane de Souza***

"ECONOMIA E NATUREZA: PADRÕES DE USO E COBERTURA DA TERRA ASSOCIADOS A ATIVIDADES AGROPECUÁRIAS E EXTRATIVISTAS DE COMUNIDADES DO SUDOESTE DO PARÁ"

Aprovado (a) pela Banca Examinadora em cumprimento ao requisito exigido para obtenção do Título de ***Mestre*** em

Sensoriamento Remoto

Dra. Maria Isabel Sobral Escada

Presidente / Orientador(a) / INPE / SJCampos - SP

Dr. Antonio Miguel Vieira Monteiro

Orientador(a) / INPE / SJCampos - SP

Dra. Silvana Amaral Kampel

Membro da Banca / INPE / SJCampos - SP

Dra. Evlyn Márcia Leão de Moraes Novo

Membro da Banca / INPE / SJCampos - SP

Dra. Jussara de Oliveira Ortiz

Membro da Banca / INPE / São José dos Campos - SP

Dr. Danilo Araújo Fernandes

Convidado(a) / UFPA / Belém - PA

Este trabalho foi aprovado por:

( ) maioria simples

(X) unanimidade

São José dos Campos, 24 de Agosto de 2016