Int J Appl Earth Obs Geoinformation 82 (2019) 101872

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

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

Detailed agricultural land classification in the Brazilian cerrado based on phenological information from dense satellite image time series

Hugo do Nascimento Bendini[a,*], Leila Maria Garcia Fonseca[a], Marcel Schwieder[b], Thales Sehn Körting[a], Philippe Rufin[b,c], Ieda Del Arco Sanches[a], Pedro J. Leitão[b,d], Patrick Hostert[b,c]

[a] National Institute for Space Research (INPE), Brazil
[b] Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, Berlin, 10099, Germany
[c] Integrative Research Institute on Transformations of Human-Environment Systems (IRI THESys), Humboldt University zu Berlin, Unter den Linden 6, Berlin, 10099, Germany
[d] Department Landscape Ecology and Environmental Systems Analysis, Technische Universität Braunschweig, Langer Kamp 19c, 38106, Braunschweig, Germany

ARTICLE INFO

Keywords:
Big data
Time-Series mining
Random forest algorithm
Land use and Land cover mapping (LULC)
Multi-Sensor

ABSTRACT

The paradox between environmental conservation and economic development is a challenge for Brazil, where there is a complex and dynamic agricultural scenario. This reinforces the need for effective methods for the detailed mapping of agriculture. In this work, we employed land surface phenological metrics derived from dense satellite image time series to classify agricultural land in the Cerrado biome. We used all available Landsat images between April 2013 and April 2017, applying a weighted ensemble of Radial Basis Function (RBF) convolution filters as a kernel smoother to fill data gaps such as cloud cover and Scan Line Corrector (SLC)-off data. Through this approach, we created a dense Enhanced Vegetation Index (EVI) data cube with an 8-day temporal resolution and derived phenometrics for a Random Forest (RF) classification. We used a hierarchical classification with four levels, from land cover to crop rotation classes. Most of the classes showed accuracies higher than 90%. Single crop and Non-commercial crop classes presented lower accuracies. However, we showed that phenometrics derived from dense Landsat-like image time series, in a hierarchical classification scheme, has a great potential for detailed agricultural mapping. The results are promising and show that the method is consistent and robust, being applicable to mapping agricultural land throughout the entire Cerrado.

1. Introduction

Brazilian agribusiness has great economic importance for the country and it is considered as one of the most relevant agribusiness frontier in the world (Spera, 2017; Ministry of Agriculture, 2016). With over half of Brazil's agricultural land falling within Cerrado biome, land clearing in this region has been central to the development of the agricultural sector (Spera, 2017). The Cerrado is the second-largest biome in Brazil (Ministry of Environment, 2019), considered as a biodiversity hotspot, providing environmental services of global importance. Despite that, the Cerrado has lost 88 Mha (46%) of its native vegetation with a projection that 31-34% of remaining biome is likely to be cleared by 2050 (Strassburg et al., 2017). This scenario underlines the need for methods that accurately map the distribution of agricultural areas and its evolution over time, so that we can ensure sustainable agricultural development and the preservation of the biome.

TerraClass is the main official LULC mapping initiative in Brazil, initially focusing on Legal Amazon deforested areas (Almeida et al., 2016) and later Cerrado (INPE, 2017). Recently, MapBiomas project (Azevedo et al., 2018) has proposed to carry out the annual fully automatic mapping of all Brazilian biomes using Landsat imagery. However, agriculture in particular is mapped at low level of thematic detail, consisting in a broad class for annual agriculture, which does not take into account the cropping practices and rotation systems.

Remote sensing data is a valuable tool for agricultural mapping, once can capture the seasonal behavior of vegetation (Wardlow et al., 2007). Coarse spatial resolution sensors (250 m – 1 km) with up to daily

* Corresponding author at: Astronautas Avenue, 1758 - 12227-010, São José dos Campos, SP, Brazil.
E-mail addresses: hugo.bendini@inpe.br (H. do Nascimento Bendini), leila.fonseca@inpe.br (L.M. Garcia Fonseca), marcel.schwieder@geo.hu-berlin.de (M. Schwieder), thales.korting@inpe.br (T. Sehn Körting), philippe.rufin@geo.hu-berlin.de (P. Rufin), ieda.sanches@inpe.br (I. Del Arco Sanches), p.leitao@geo.hu-berlin.de (P.J. Leitão), patrick.hostert@geo.hu-berlin.de (P. Hostert).

https://doi.org/10.1016/j.jag.2019.05.005
Received 29 December 2018; Received in revised form 4 May 2019; Accepted 8 May 2019
Available online 18 June 2019

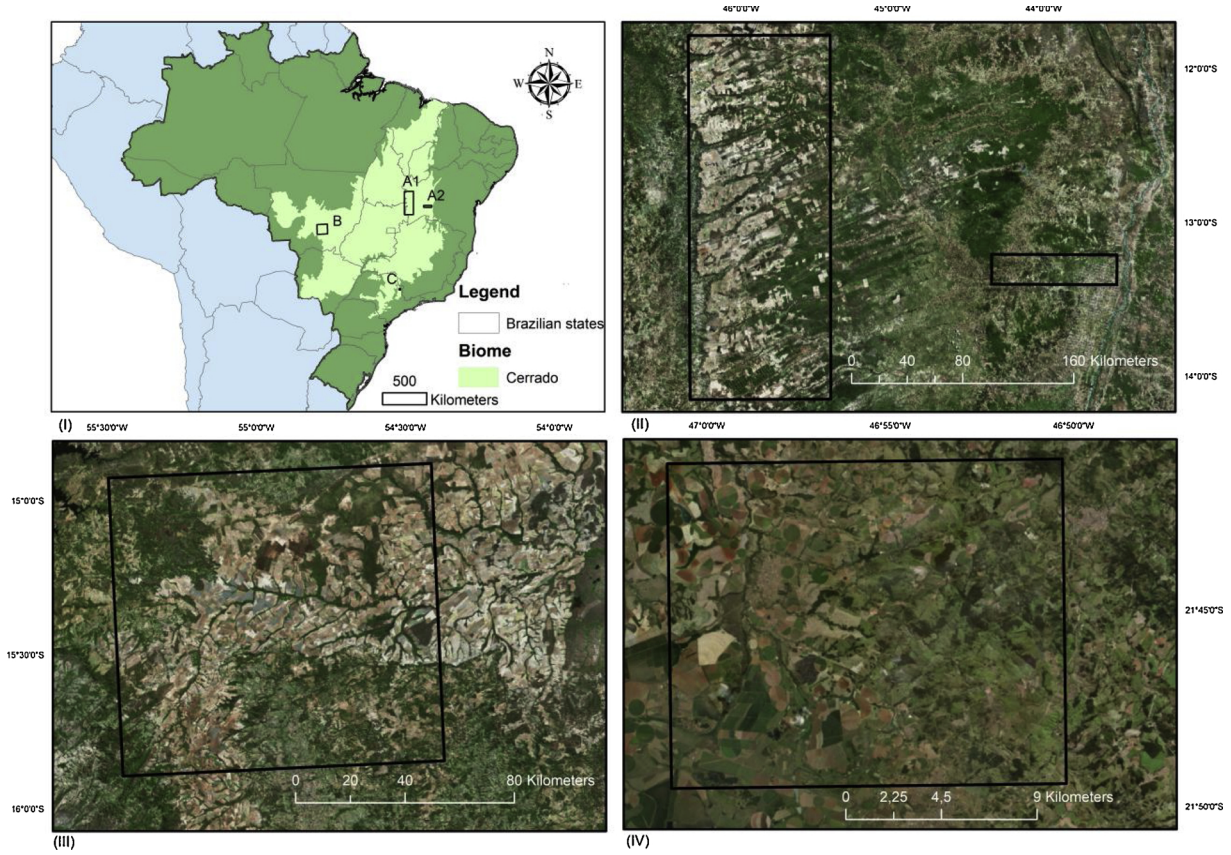

Fig. 1. Locations of the study sites in the Cerrado biome (I). The sites A1 and A2 (II) are located in western Bahia, site B in southeastern Mato Grosso (III) and site C in northeastern São Paulo (IV).

revisit times have been widely used for agricultural mapping (Sakamoto et al., 2005; Wardlow et al., 2007) and several studies highlighted benefits of time series for agricultural mapping in Brazil (Esquerdo et al., 2011; Rudorff et al., 2010; Arvor et al., 2011; Borges and Sano, 2014; Brown et al., 2013). However, even though their spatial resolution has been shown to be sufficient for mapping large-scale agricultural practices, such as double and single cropping systems, it does not allow for the detection of smaller fields, due to the spectral mixture of the different targets, and which includes crops that has significant importance for familiar agriculture.

With advances in computational processing performance and the development of consistent methods to combine data from different sensor there is an increase in advanced LULC mapping approaches that use Landsat-like images (Zheng et al., 2015; Peña and Brenning, 2015; Pan et al., 2015; Wulder et al., 2012; Müller et al., 2015; Rufin et al., 2015; Schwieder et al., 2016; Bendini et al., 2016; Bendini et al., 2017), which can overcome the spatial resolution limitation. However, temporal resolution is still a problem, since in many cases, especially in tropical regions, there is a high cloud cover and the variability in the agricultural calendar is highly dynamic, with different crop seasons within the same year and different cultivation practices (e.g., no tillage, center pivot irrigation, crop-livestock integration and the use of early varieties). The 16-days temporal resolution may not be sufficient (Bendini et al., 2017). Combining different sensors is an alternative to increase the temporal resolution, especially after the launch of Sentinel-2B, that provide with Sentinel-2A and Landsat-8 three medium resolution sun synchronous satellites on orbit. However, to map also the past before the Sentinel launch in 2013, in this spatial resolution, it is necessary to consider Landsat TM and ETM+, and consequently deal with the Scan Line Corrector (SLC)-off data (Schwieder et al., 2016).

The use of phenological metrics (phenometrics) extracted from image time series can be an important strategy for the development of agriculture mapping methods. Different phenometrics were explored for this purpose, mainly using MODIS time series (Sakamoto et al., 2005; Arvor et al., 2011; Borges and Sano, 2014; Körting et al., 2013) and recently also with Landsat-like images (Pan et al., 2015; Müller et al., 2015; Bendini et al., 2016; Bendini et al., 2017; Schmidt et al., 2016). However, the full potential of these metrics has not yet been fully explored, considering their applicability for mapping multiple cropping systems considering the different cultivation practices in highly dynamic agricultural regions.

Griffiths et al. (2019) reached overall accuracy of 81% mapping 12 agricultural classes in Germany using 10-day interval composites of Sentinel 2A and Landsat imagery (Griffiths et al., 2019). However they did not explore the use of phenometrics showing that it could improve their results, which is linked to the need of more sophisticated gap-filling techniques. Rufin et al. (2019) showed the potential of the Radial Basis Function (RBF) fitting to derive gap-filled image time series to classify cropping practices in Turkey using RandomForest algorithm, achieving accuracies above 90% (Rufin et al., 2019). They opened a perspective about the potential of phenological metrics as a way for integrating features from multiple crop seasons which can be important for mapping agriculture with higher thematic detail.

Hierarchical classification approaches is an efficient strategy to deal with classification problems when evolving a high number of classes, being possible to achieve higher accuracies in comparison with classical approach (Lebourgeois et al., 2017). Furthermore, in the agricultural domain, hierarchical approach enables the evaluation of the cropping practices and crop types independently of the use of a *priori* cropland mask.

The combination of phenological information derived from dense Landsat image time series combined to the use of hierarchical classification approaches and non-parametric classifiers such as RF, for our knowledge was still not applied. Thus, the objective of this work is to

use phenometrics derived from dense satellite image time series for classifying agricultural land in the Cerrado biome at different thematic levels with a hierarchical approach and RF classification.

2. Materials and methods

2.1. Study sites and definitions for the hierarchical classification

We tested the method in three different study areas within the Cerrado (Fig. 1). The study areas were in the west of Bahia State (regions A1 and A2), southeast of Mato Grosso (region B) and northeast of São Paulo (region C). Despite all of these areas being in the Cerrado biome, they differ with respect to climate conditions, vegetation and agricultural practices.

The areas in western Bahia and southeastern Mato Grosso (Fig. 1; sites A1, A2, and B) consist mostly of large-scale and market-oriented agriculture. The region of western Bahia belongs to the most recent agricultural frontier of the Cerrado, which is called the Matopiba region (Miranda et al., 2014), a continuous zone formed by the states of Maranhão, Tocantins, Piauí, and Bahia. The study area in southeastern Mato Grosso is characterized by intensive double-cropping rotations, while western Bahia features mostly single-cropping regimes. The study area in northeastern São Paulo (Fig. 1; region C) is considered a smallholder agricultural zone, characterized by high intra and inter-field spatial variability. We collected data at different thematic detail (Fig. 2) to test the approach using hierarchical classification scheme. Field surveys were conducted during the 2015–2016 cropping seasons. For the study area in western Bahia, we conducted the survey with the "Rally da Safra 2016″ team (AGROCONSULT, 2018), which is a project that visits several farms on Brazil to evaluate crop yield during the growing peak. We registered the samples locations for each class using a GPS and collected information about crop rotations with farmers. The fieldwork protocol and the database in region B are available in (Sanches et al., 2018). We used the same protocol for region C. We used TerraClass maps and interpretation of Google Earth imagery to collect additional samples for non-cropped classes. Field boundaries were digitized to obtain polygon database. We generated 841 polygons, where a small fraction of randomly pixels were sampled from each polygon in order to incorporate intraclass variability, totalizing 40,385 samples.

We used a hierarchical classification approach by which level 1 classes domains are isolated, and land cover is classified by correspondence at level 2. Then we isolated the pixels classified as "Annual crop" and used it to classify land use by correspondence for each domain for the subsequent nomenclature levels.

The land cover classes (Level 2) are based on the nomenclature of the Systematic Survey of Agricultural Production of the Brazilian Institute of Geography and Statistics (IBGE, 2016). The natural vegetation classes consist into three main Cerrado physiognomies, based on the definition of TerraClass Cerrado (INPE, 2017): forest, savanna and natural grasslands (Ribeiro and Walter, 2008). The Crop Group (Level 3) is defined by the main agricultural practices in the Cerrado region and the Crop Rotation level (Level 4) consists in the most detailed level, with crop rotation types definitions.

2.2. Satellite data

We used all available ETM + and OLI data for the study areas (Path/Row 226/070, 225/070, 226/071, 225/071, 220/068, 220/069, 220/070, 219/069 and 219/075), acquired between April 2013 and April 2017. Assuming an 8-day temporal resolution, this 4-year period contains 186 potential observations. The images were obtained from the US Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA). These data are provided with level 1 geometric correction (L1TP). Landsat 7 imagery was converted to surface reflectance by the atmospheric correction algorithm LEDAPS (Landsat Ecosystem Disturbance Adaptive Processing) (Masek et al., 2006), and Landsat 8 data were corrected using LaSRC (U.S. Geological Survey, 2017; Vermote et al., 2016). We used the Enhanced Vegetation Index (EVI) (Liu and Huete, 1995), which is known to increase sensitivity for biomass estimation through a de-coupling of the canopy background from the signal and a reduction in atmospheric and soil reflectance influence (Huete et al., 2002).

2.3. Landsat dense time series

Limiting factors of a dense time series are sensor errors and cloud cover. To overcome these constrains, Schwieder et al. (2016) used a weighted ensemble of Radial (Gaussian) Basis Function (RBF) convolution filters to approximate the missing data in a Landsat time series. To approximate the given EVI observations into dense 8-day time series without data gaps, we used the RBF approach (Schwieder et al., 2016) with some adaptations. Let $f(t)$ be a time series, where $t \in \{1, ...,N\}$. The approximated values $y(t)$ are calculated by Eq. 1.

$$y(t) = \frac{\sum_{i=1}^{T} [f(t) \otimes \boldsymbol{K}_i]\boldsymbol{W}_i}{\sum_{i=1}^{T} \boldsymbol{W}_i} \quad (1)$$

where $\otimes$ is an operator for convolution, the kernels for the convolution ($\boldsymbol{K}_i$) are given by the Gaussian function (Eq. 2), and T is the total number of standard deviations (σ) that can be used for the kernel calculation.

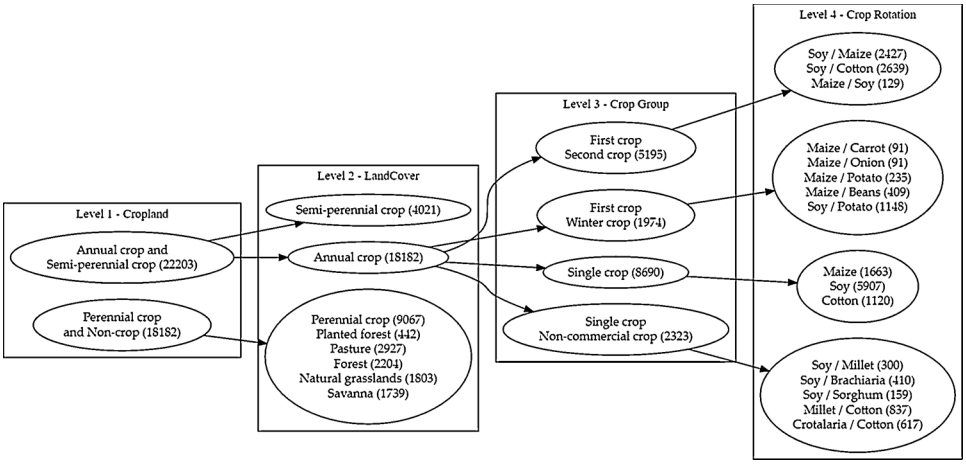

Fig. 2. Thematic levels for definition of hierarchical classification.

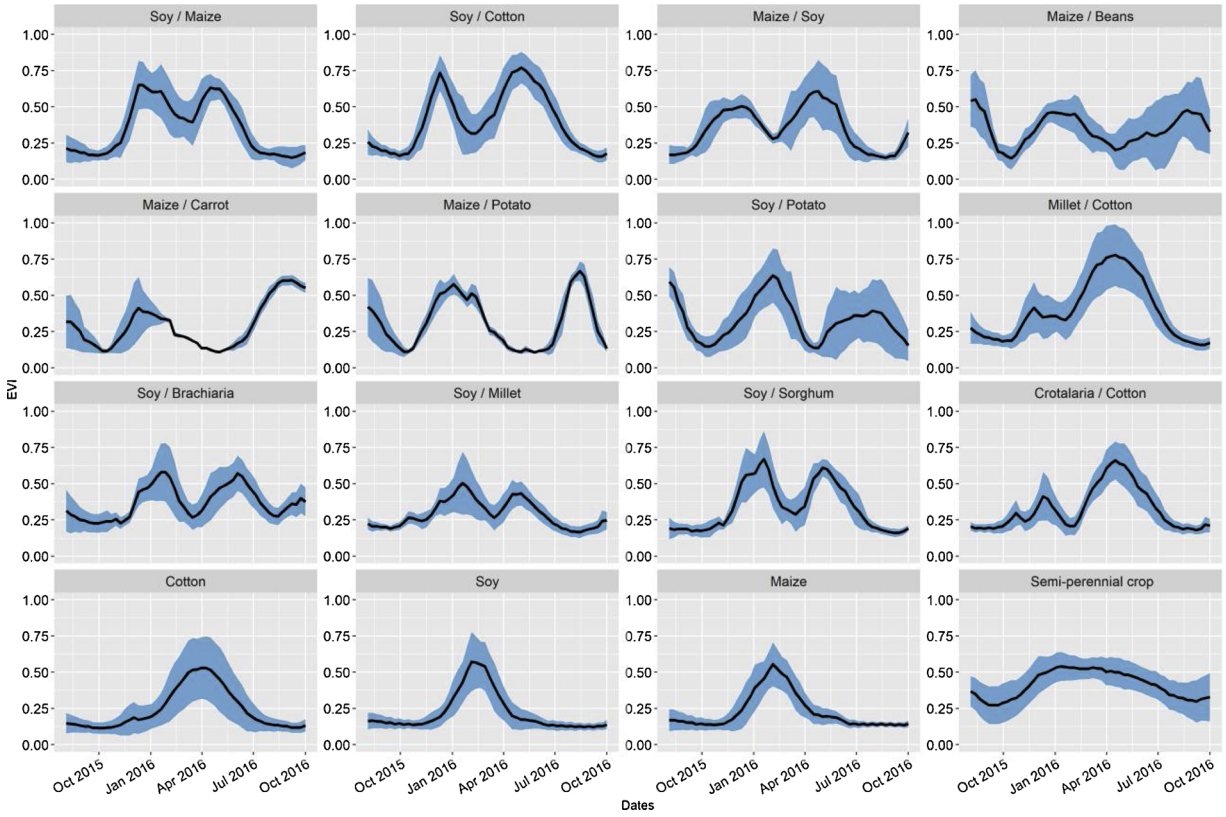

Fig. 3. Averaged EVI phenological profiles for each agriculture class of level 4 in the season 2015–2016 (black lines), with their respective standard deviations (blue margins) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

$$\boldsymbol{K}_i = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_i}\right)^2} x \in \{1, \ldots L\} \quad (2)$$

The size of the kernel window (L) is given by σ, which directly translates into the number of observations (at 8-day intervals). The total kernel width is limited to the points in time, which delineate 90% of the area under the Gaussian kernel. Increasing σ expands the kernel window and lowers the kernel values. We used three different kernels with $\sigma_1 = 0.5$, $\sigma_2 = 1.0$, $\sigma_3 = 3.0$. The final approximation is the weighted average of the results of the three temporal convolution filters, where the weights (W_i) are also calculated by the kernel convolution (Eq. 3), but applied in $d(t)$, which is a vector that expresses data availability (Eq. 4), where cloud contaminated pixels were masked by the Fmask algorithm (Zhu et al., 2015). The more data available in each kernel window relative to the total kernel window, the higher its weights in the final aggregation of the different kernels.

$$\boldsymbol{W}_i = d(t) \otimes \boldsymbol{K}_i \quad (3)$$

where,

$$d(t)\begin{cases} 0, & if\ f(t) = NA \\ 1, & if\ f(t) \neq NA \end{cases} \quad (4)$$

We did not use a *priori* outlier detection. After an expert-driven visual inspection, we observed that using outlier detection masked some intrinsic variations in the phenological profiles of agricultural targets (abrupt greening, induced senescence or harvesting). Previous studies showed that approaches using cloud masking with interpolation methods are promising in reducing noise for this purpose in similar regions (Bendini et al., 2017). Rufin et al., 2019 showed the potential of RBF for deriving intra-annual Landsat dense time series for mapping cropping practices in Turkey, where they reported an average of around 20.42 of clear sky observations (CSOs) per pixel during the study period (Rufin et al., 2019).

2.4. Phenometrics

We obtained the phenological parameters using TIMESAT v3.2 software (Jönsson and Eklundh, 2015), where seasonal data are extracted from the time series for each growing season of the focal year. In Brazil, the agricultural year of most crops is defined as between August of a given year and October of the following year; therefore, when extracting seasonality parameters, we used the period of August 4, 2015, to October 1, 2016, as the focal year. We fitted the time series using the Savitzky-Golay filter (Jönsson and Eklundh, 2004) with a window size of 4. A set of 13 phenometrics were derived for each season (S1 and S2). Parameters included day-of-the-year (DOY) of start, mid, end, and length of season and phenological proxies like peak and base value, seasonal amplitude or rate of increase and decrease (see the full list of predictors in Table B4). We also used the polar features, which the purpose is represent time series by projecting values onto angles in the interval $[0,2\pi]$. So that, we can obtain the coordinates of a closed shape and calculate the area of the resulting shape for each of the quadrants ($[\pi, 3\pi/2]$, $[\pi/2, \pi]$, $[0, \pi/2]$ and $[3\pi/2, 2\pi]$), which are supposed to represent the seasons. More details can be found in Körting et al. (2013).

2.5. Random forest classification

After the time series feature extraction, we used our field database to train RF (Breiman, 2001) and obtained a classifier for each nomenclature level (Fig. 2). RF needs two parameters to be tuned including the number of trees (ntree), and the number of variables (mtry). The mtry was empirically set to 5, and ntree values of each RF classification model after tuning were respectively 50, 50, 70 and 90. The final maps were validated using exhaustive method based on Monte Carlo simulation (Rubinstein and Kroese, 2008) where, for each model, 1000 simulations were carried out by randomly selecting 70% of samples to

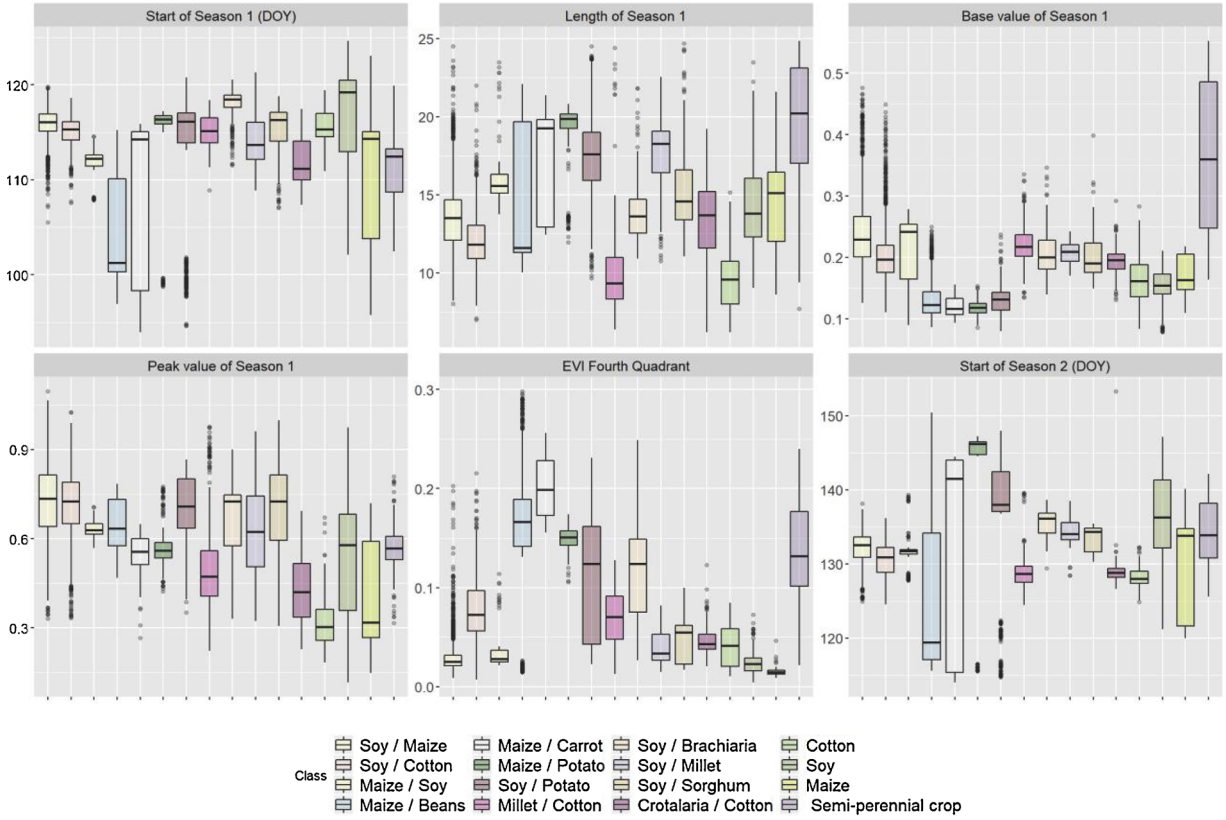

Fig. 4. Boxplots with the mean, 25, 75 percentiles and outliers of EVI phenological parameters for the main classes of level 4 classification.

train and 30% for validation. For each subdivision, a confusion matrix was calculated, and the average confusion matrix was used to derive the overall accuracy (Chinchor and Sundheim, 1993) and the class f1-scores (Shapiro, 1999) for each model. The "randomForest" package in R was used for our classification tasks (R Development Core Team, 2017; Liaw and Wiener, 2002). For the visualization of the final maps, we grouped the different non-crop classes into one. We used the Modified Normalized Difference Water Index (MNDWI) (Xu, 2005) to derive a mask of the water bodies.

3. Results and discussions

3.1. EVI temporal profiles

We derived pixel-wise EVI fitted time series with 8-day temporal resolution for the period of 2013–2017. Fig. 3 shows averaged profiles for the target period of 2015–2016 (correspondent to the field work period); with the respective standard deviations for each agriculture class of level 4 based on the sampled pixels of all the study sites (see Fig. A1 and A2 in Appendix for the level 2 and 3).

By assessing the Fmask products for the same period we found an average of 30.45 CSOs per pixel (minimum 0; maximum 46) (see Figure A3 in Appendix for spatially explicit visualization of the CSOs count and Figure A4 for an example of RBF interpolation in a selected pixel profile).

The temporal profiles of the First crop / Second crop classes are mostly represented by a first season of soy, or maize, generally planted from late September to mid-November or even in early December. The harvest period occurs from late January to mid-March. The second season commonly consists of maize or cotton and planting occurs between late February and early March, depending on when the first-season crops are harvested, generally from June to early July. Our approach sought to assesses agricultural phenological patterns not only for major crops such as maize, soy and cotton, but also minor crops which are significant at smaller scale. Nonetheless, we observed that our phenological profiles for the major crops agree with findings of other authors (Arvor et al., 2011; Oliveira et al., 2014), which worked with coarse spatial resolution data and focused only on larger areas of Mato Grosso state.

The first season of First crop / Winter crop classes begin approximately one month later than First crop / Second crop class. In region C, winter crops are represented by potatoes or minor crops like carrots; these are planted from mid-March to June and harvested by late September or early October. The slope of the green-up curve is subtly less pronounced, because the first crop in this class is usually maize, which has longer green-up period than soy (Nguy-Robertson et al., 2012). The EVI values in Soy / Maize and Soy / Cotton off-seasons are higher than when second season is potato. Rotations systems in Brazil are usually based on no-tillage, unless when the first season is potato, once this crop require tillage before planting. This may explain the low EVI values.

Non-commercial crops are planted to maintain a constant vegetation cover during the off-season when no-tillage is used. Millet and crotalaria presented a very short cycle with low EVI peaks, and are usually planted right after or before main season. We observed this affected the TIMESAT season detection, and phenometrics extraction. Soy / Sorghum and Soy / Maize spectral temporal behavior are similar. Maize and sorghum plants have similar erectophile structure, what can lead to confusion. This was also observed by Arvor et al. (2011) and Zheng et al. (2015) reported the same issue with wheat and barley in Central Arizona (Arvor et al., 2011; Zheng et al., 2015). The second season in Maize / Beans temporal profiles behaved differently than expected once we cannot see clearly second season. Soy and common beans cultivated in Brazil are also similar in terms of structure, but common beans have shorter cycle, being mostly cultivated in family farming systems and irrigated areas. Planting dates can vary widely, and combined to the short cycle reflected in problems with the phenological characterization by EVI fitted values. This point to the fact

that cloud cover is also contributing to mask important phenological phenomena in such dynamic agricultural systems, what evidence the potential of SAR sensors, in order to provide weather-independent observations.

Single-crop class is represented by cotton, soy and maize. When cultivated in a single-cropping system, cotton is planted between October and December and harvested from April to July. Single crops are usually planted with soil tillage, so the EVI values on off-season are lower than for double cropping systems.

3.2. Phenometrics

We used the derived EVI temporal profiles on TIMESAT to calculate pixel-wise phenological parameters for the season 2015–2016. The distribution of these parameters in each class highlighted phenological differences. Fig. 4 displays the distribution of selected EVI phenological parameters for classes in level 4.

Detecting planting date is an important challenge for remote sensing in agriculture (Sakamoto et al., 2005), and can be applied for many applications such as risk assessment for rural credit and yield estimation. Analyzing the start of seasons, which is related to planting date, we can observe that phenometrics are correctly describing the studied crops. Soy planting dates varied between late September and the end of November. Late planting dates in this case are observed especially when soy is planted as single crop or in rotation with non-commercial systems. Usually, in double-cropping systems farmers use early varieties of soy, in order to have time to plant the second crop season still during the rainy periods (Cattelan and Dall'Agnol, 2018). This might be reason for higher values of length of season for soy in single cropping systems or with non-commercial crops.

Our phenometrics showed the potential for description of crop phenology in the Cerrado, which is important from an agronomic perspective, for example, base values for winter and single crops are considerably lower than other crops, which are possibly related to crops as potatoes, that require intensive tillage operations for planting, while other double-cropping systems are planted mostly without tillage. The same is observed for single crops, largely represented by soy. Single crops are usually planted through conventional systems, i.e. soil tillage. This highlights the potential for using this information on the discrimination of no-tillage agricultural areas, which is important for assessments of long-term carbon dynamics in agricultural lands (Foley et al., 2011).

Sugarcane crops constitute the Semi-perennial class; thus, higher values were expected for length of season, since a normal sugarcane cycle is about 9–18 months. By integrating different varieties of sugarcane in our samples, the phenometrics showed start, end, and length of season for this class varying widely. Despite this, variables generated with polar representation, which does not depend on season detection, and hierarchical classification efficiently separated semi-perennial crops from annual crops. This can be also observed for the First crop / Winter crop classes, once their EVI values are higher in the 4th quadrant, which represent the period from early August to mid-November and June to October. The average season of the winter crops extend from April to October, when the other classes are either still being planted, greening-up or in senescence and being harvested.

3.3. Agricultural land classification

Table 1 shows the overall accuracy and the class f1-scores (see Tables B1, B2, B3 and B5 in Appendix for the complete confusion matrices of each classification level and Fig. A5 shows boxplots of overall accuracies of each model).

As we can see in Table 1 lower accuracies were found for classes Soy/Millet and Soy/Sorghum. Fig. 5 displays predicted versus the observed values for each class in level 4, using the model that achieved the highest overall accuracy in the Monte Carlo simulation, in order to visualize accuracy and misclassifications.

Fig. 5 shows there was misclassification between these classes with the Soy/Maize class. This can be explained by the fact of maize, millet and sorghum plants have similar characteristics, as we observed by the phenological profiles (Fig. 3). By analyzing confusion matrix, major confusion is found between Soy / Sorghum and Soy / Maize, Soy / Maize and Soy / Cotton. We can also observe misclassification between Maize and Soy and Soy / Cotton with Crotalaria / Cotton. This misclassification problem was also reported by other authors (Arvor et al., 2011; Picoli et al., 2017) and associated to spectral similarity between classes. This shows opportunities on the use of methodologies that explore different spectral bands, such as the ones in red-edge domain as provided by Sentinel-2.

The cropping patterns provide information about the spatial distribution of the croplands. Fig. 6 presents maps of regions A1 and A2 obtained at level 2 and 4 of hierarchical classification. The water bodies' mask revealed some small dams, most of which are close to center-pivot irrigated areas (see insets in Fig. 6).

The map of Fig. 6 showed that although single-cropping system is not recommended, is still being widely used in this region, which agrees with the results of Spera et al. (2016) which showed that, in 2015, 85% of the large-scale agriculture in the Matopiba region was based on single-cropping systems (Spera et al., 2016). The soils in this region are sandier and less physically suitable than Mato Grosso, and the region receives less rainfall; these aspects pose a challenge to adopting double-cropping regimes (Spera, 2017). Double-cropping systems occur mostly in irrigated areas or in the western side of the region A1, where precipitation is higher (Dourado et al., 2013). Perennial crops were mapped in the northern study area, especially in irrigated areas. An area in the eastern part of region A2 also has a large concentration of perennial crops, which may be related to this region's proximity to the São Francisco River, making it better for irrigation.

Fig. 7 presents map of region B (southeast of Mato Grosso) obtained at levels 2 and 4 of the hierarchical classification. The map shows most of the agricultural areas based on double-cropping systems of soy/maize and soy/cotton, but some small areas of single cropping and non-commercial crops remain, mostly in the northeast. We can also observe semi-perennial crops in the south, which includes the municipality of Jaciara, one of the biggest sugarcane producers of Mato Grosso State (IBGE, 2016). This demonstrates the capacity of the proposed method to map semi-perennial crops.

Fig. 8 presents the map of region C (northeast of São Paulo), obtained at levels 2 and 4. This map shows a great heterogeneity of classes, because this is a smallholder agricultural zone, where farmers grow different crops throughout the year.

Most of the agricultural areas are concentrated on western side of the study area, which belongs to Cerrado biome and is characterized by flat topography. This is one scenario of intensive agriculture, with double-cropping systems using center-pivot irrigation, and different crop systems are adopted within the same area. The map of Fig. 8 shows what we expected by the field work, where we observed that there was large areas of sugarcane and Soy/Potato crop rotations. Although the high accuracies for the "Perennial crop" class, we could observe visually that there was some misclassification errors where natural vegetation were included. In general, we could observe that these misclassified areas are located mostly near areas of transition to different biomes (Caatinga and Atlantic Forest).

4. Conclusions and outlooks

The presented results show that phenometrics derived from dense Landsat image time series can be used for describing phenological patterns of complex and dynamic agricultural environment on the Brazilian Cerrado and be used to map the distribution of crops over these areas with accuracies above 90%. To our knowledge, this is the first paper to provide a method for mapping agricultural land in the

Table 1
Overall accuracy and class wise f1-scores (within parentheses) obtained for the different hierarchical classification levels.

L1 (0.985)*	L2 (0.995 and 0.968)**	L3 (0.977)*	L4 (0.956)*
Annual crop and Semi-perennial crop (0.988)	Annual crop (0.996)	First crop / Second crop (0.971)	Soy / Maize (0.940) Soy / Cotton (0.952) Maize / Soy (0.993)
		First crop / Winter crop (0.998)	Maize / Carrot (0.994) Maize / Onion (0.967) Maize / Potato (0.997) Maize / Beans (0.999) Soy / Potato (0.9984)
		Single crop (0.991)	Maize (0.9224) Soy (0.976) Cotton (0.946)
		Single crop / Non-commercial crop (0.924)	Soy / Millet (0.885) Soy / Brachiaria (0.934) Soy / Sorghum (0.827) Millet / Cotton (0.933) Crotalaria / Cotton (0.912)
Perennial crop and Non-crop (0.983)	Semi-perennial crop (0.985) Perennial crop (0.992) Planted forest (0.893) Pasture (0.969) Forest (0.938) Natural grasslands (0.939) Savanna (0.924)		

* The overall accuracy of the model for the respective classification level.
** The overall accuracies of the model for the classification level of the respective hierarchical classes.

Brazilian Cerrado using Landsat-like data to consistently and reproducibly achieve a detailed distinction between crops. This framework can also be tested by integrating other satellite data such as Sentinel 2-A and 2-B.

We also demonstrated the potential application of phenometrics to describe phenological information about the major crops that Brazilian farmers grow in the Cerrado biome. This provides information that can be useful for understanding agricultural practices and improve insights for methods for detecting planting date and on the discrimination of no-tillage agricultural areas, which is important for assessments of long-term carbon dynamics in agricultural lands and understanding the effects of policy, trade, and global and technological change on food security.

The classification results showed that the classes of Soy / Sorghum

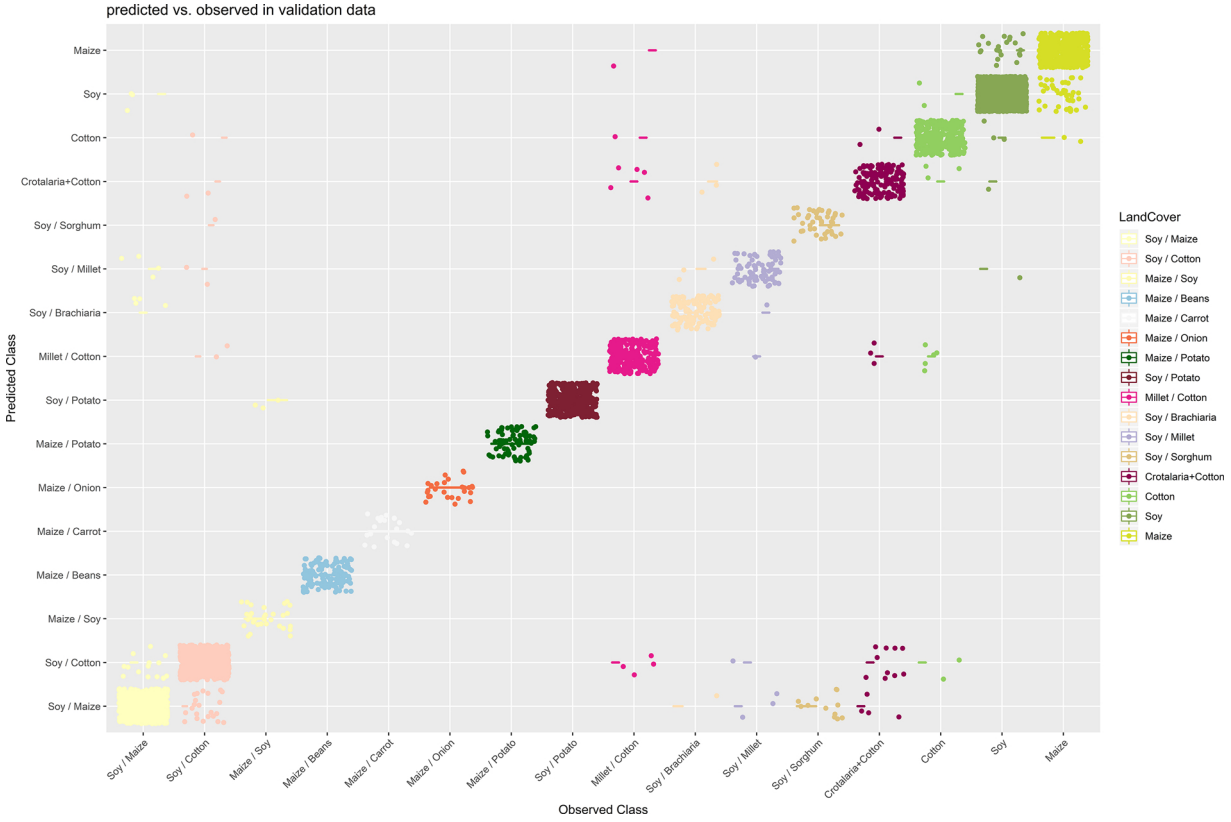

Fig. 5. Predicted versus the observed values for each class in level 4, based on the results obtained by the best model.

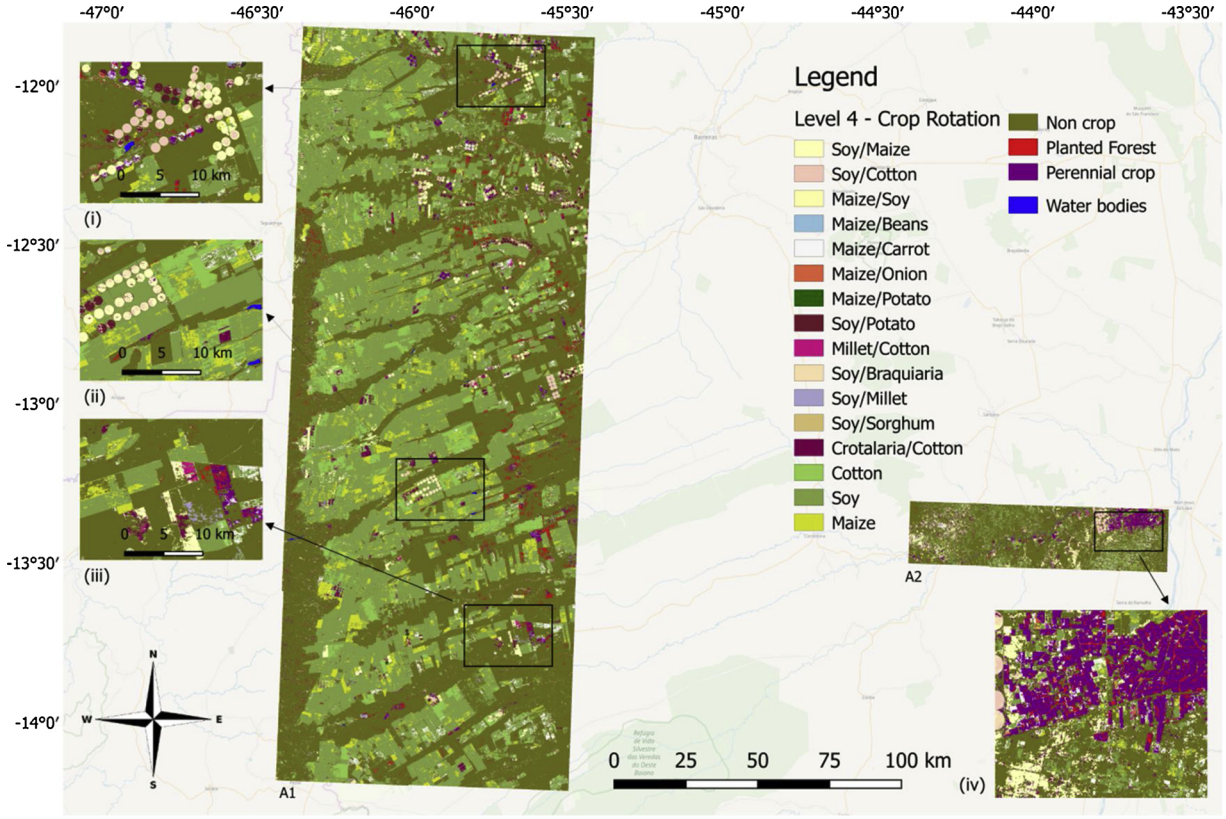

Fig. 6. Map of regions A1 and A2, obtained at level 2 and level 4 of the hierarchical classification approach. The insets (i), (ii), (iii) and (iv) show zoomed-in areas of the maps.

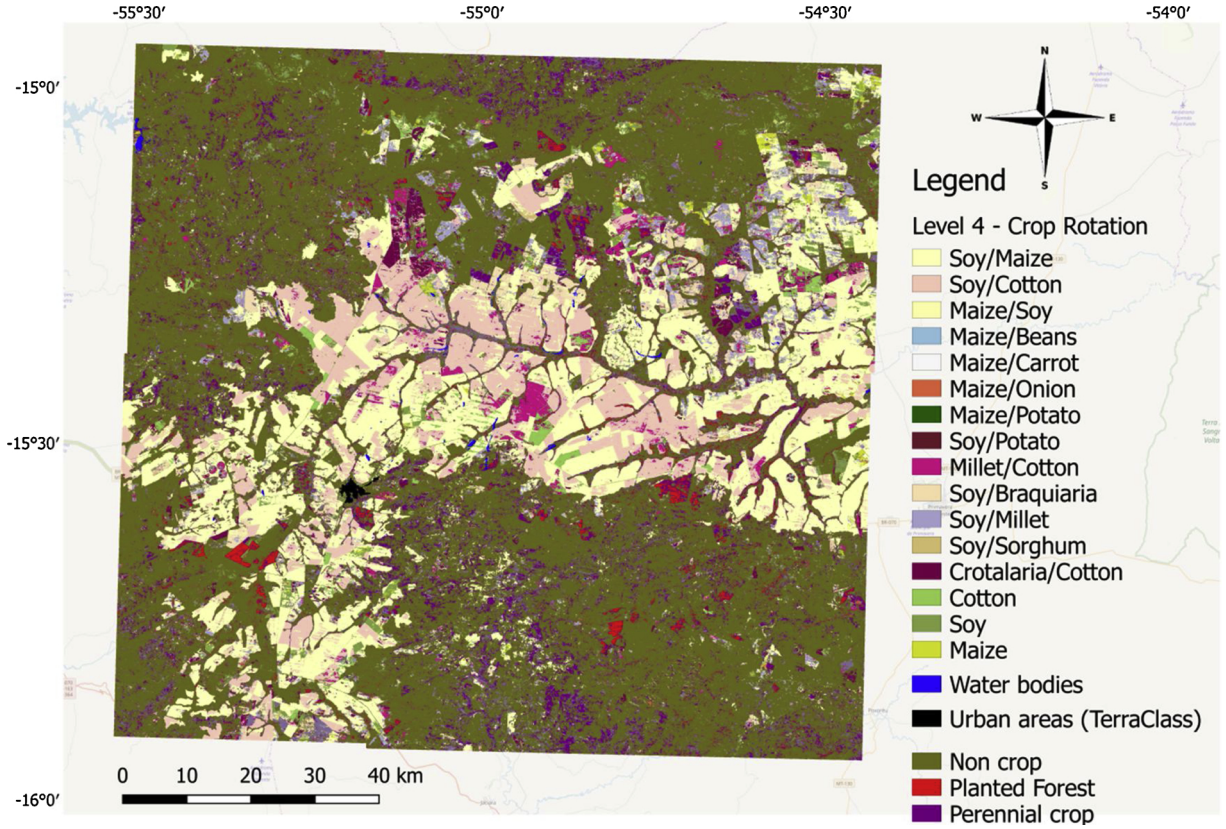

Fig. 7. Map of region B, obtained at level 2 and 4 of the hierarchical classification approach.

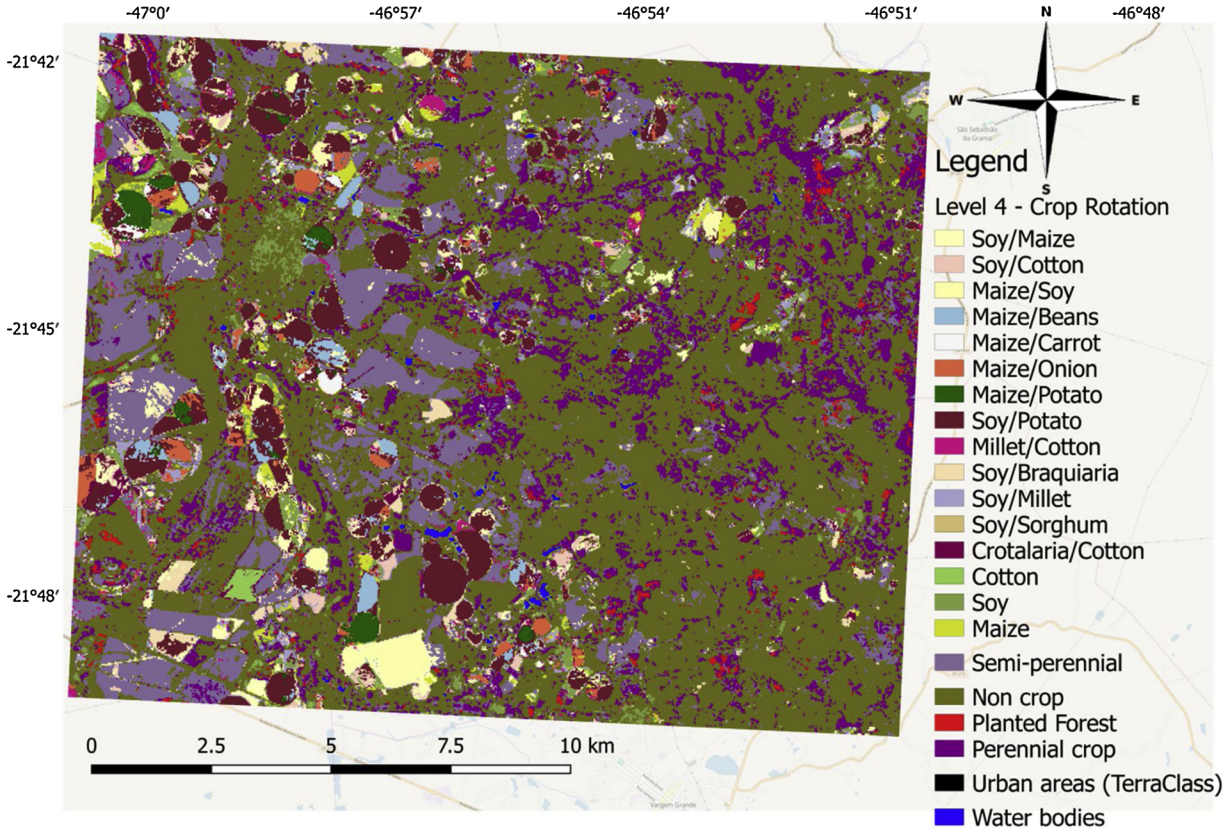

Fig. 8. Map of region C, obtained at level 2 and 4 of the hierarchical classification.

and Soy / Maize still presented some confusion due to their spectral temporal similarity, showing the need of more studies about methodologies that explore different spectral bands, such as the ones in the red-edge domain as provided by Sentinel-2 sensor.

A narrow interval between seasons is observed in rotation systems with shorter cycle crops like common beans and non-commercial crops such crotalaria and millet and resulted in challenging for the phenological characterization and classification. But more studies about the impact of cloud covering in specific periods and how it is impacting the fitting are necessary in these regions. This also highlights the potential of SAR sensors, which are weather-independent and can enhance the classification results for these crops. Variables generated with polar representation, which does not depend on season detection, showed great potential for classifying winter crops and semi-perennial crops when combined to hierarchical classification.

For future works, we aim to apply this method to the entire Brazilian Cerrado biome for long-term agricultural analysis in a property level.

Acknowledgments

This study was financed in part by the "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)" – Finance Code 001. The authors acknowledge the "Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)" e-science program (grant 2014-08398-6) for the financial support of the field visits, the CAPES "Doctoral Sandwich Program Abroad" for the scholarship during the stay at Germany and the project "Development of systems to prevent forest fires and monitor vegetation cover in the Brazilian Cerrado" (World Bank Project #P143185) – Forest Investment Program (FIP). It further contributes to the Global Land Program (https://glp.earth) and Landsat Science Team (http://landsat.usgs.gov/Landsat Science Team2012-2017.php). We also acknowledge Andreas Rabe for RBF code programming.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jag.2019.05.005.

References

AGROCONSULT, 2018. AGROCONSULT: Consultoria E Projetos. Available online (accessed on March. http://www.agroconsult.com.br/en/.

Almeida, C., Coutinho, A., Esquerdo, J., Adami, M., Venturieri, A., Diniz, C., Dessay, N., Durieux, L., Gomes, A., 2016. High spatial resolution land use and land cover mapping of the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data. Acta. Amazon. 46, 291–302.

Arvor, D., Jonathan, M., Meirelles, M.S.O.P., Dubreuil, V., Durieux, L., 2011. Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. Int. J. Remote Sens. 32 (22), 7847–7871.

Azevedo, T., Souza, J.C.M., Shimbo, J., Alencar, A., 2018. MapBiomas initiative: mapping annual land cover and land use changes in Brazil from 1985 to 2017. AGU Fall Meeting Abstracts.

Bendini, H.N., Sanches, I.D.A., Körting, T.S., Fonseca, L.M.G., Luiz, A.J.B., Formaggio, A.R., 2016. Using Landsat 8 Image Time Series for Crop Mapping in A Region of Cerrado, Brazil. ISPRS. J. Photogramm Remote. Sens. XLI-B8, 845–850.

Bendini, H.N., Fonseca, L.M.G., Körting, T.S., Marujo, R., Sanches, I.D.A., Arcanjo, J., 2017. 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. Brazilian Journal of Cartography 69 (5).

Borges, E.F., Sano, E.E., 2014. Séries temporais de EVI do MODIS para o mapeamento de uso e cobertura vegetal do oeste da Bahia. Bol. Ciênc. Geod 20 (3), 526–547.

Breiman, L., 2001. Mach. Learn. 45 (5), 5–32.

Brown, J.C., Kastens, J.H., Coutinho, A.C., Victoria, D.C., Bishop, C.R., 2013. Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. Remote Sens. Environ. 130, 39–50.

Cattelan, Alexandre José, Dall'Agnol, Amelio, 2018. The rapid soybean growth in Brazil. Embrapa Soja-Artigo em periódico indexado (ALICE).

Chinchor, N., Sundheim, B., 1993. MUC-5 evaluation metrics. Paper Presented at the Proceedings of the 5th Conference on Message Understanding.

Dourado, C.S., Oliveira, S.R.M., Avila, A.M.H., 2013. Análise de zonas homogêneas em séries temporais de precipitação no Estado da Bahia. Bragantia 72, 192–198.

Esquerdo, J.C.D.M., Zullo Júnior, J., Antunes, J.F.G., 2011. Use of NDVI/AVHRR time-series profiles for soybean crop monitoring in Brazil. Int. J. Remote Sens. 32 (13), 3711–3727.
Foley, J.A., Ramankutty, N., Brauman, K.A., et al., 2011. Solutions for a cultivated planet. Nature 478, 337–342.
Griffiths, Patrick, Nendel, Claas, Hostert, Patrick, 2019. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. Remote Sens. Environ. 220, 135–151.
Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213.
IBGE, 2016. IBGE: Brazilian Institute Of Geography And Statistics. Available online: (accessed on October. https://sidra.ibge.gov.br/pesquisa/pam/tabelas.
INPE, 2017. Mapeamento De Uso E Cobertura Vegetal Do Cerrado. Available online (accessed on 1 March. http://www.dpi.inpe.br/tccerrado/index.php?mais=1.
Jönsson, P., Eklundh, L., 2004. TIMESAT – a program for analyzing time-series of satellite sensor data. Comput. Geosci. 30, 833–845.
Jönsson, P., Eklundh, L., 2015. TIMESAT 3.2 With Parallel Processing Software Manual. Lund University, Sweden, pp. 22–24.
Körting, T.S., Fonseca, L.M.G., Câmara, G., 2013. GeoDMA—geographic data mining analyst. Comput. Geosci. 57, 133–145.
Lebourgeois, Valentine, et al., 2017. A combined random forest and OBIA classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated Sentinel-2 time series. VHRS and DEM)." Remote Sensing 9.3, 259.
Liaw, A., Wiener, M., 2002. Classification and regression by randomforest. R. News 2, 18–22.
Liu, H.Q., Huete, A., 1995. Feedback based modification of the NDVI to minimize canopy background and atmospheric noise. Ieee Trans. Geosci. Remote. Sens. 33 (2), 457–465.
Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., 2006. A Landsat surface reflectance dataset for North America, 1990-2000. Ieee Geosci. Remote. Sens. Lett. 3, 68–72.
Ministry of Agriculture, 2016. Livestock and Supply: Statistics Data. Available online: (accessed on October). http://www.agricultura.gov.br.
Ministry of Environment. Available online: http://www.mma.gov.br/ (accessed on October 2016).
Miranda, E.E., Magalhães, L.A., Carvalho, C.A., 2014. Proposta De Delimitação Territorial Do Matopiba. Nota Técnica 1. EMBRAPA. Grupo De Inteligência Territorial Estratégica (GITE).
Müller, H., Rufin, P., Griffiths, P., Siqueira, A.J.B., Hostert, P., 2015. Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. Remote Sens. Environ. 156, 490–499.
Nguy-Robertson, A., Gitelson, A., Peng, Y., Viña, A., Arkebauer, T., Rundquist, D., 2012. Green leaf area index estimation in maize and soybean: combining vegetation indices to achieve maximal sensitivity. Agron. J. 104 (5)), 1336–1347. https://doi.org/10.2134/agronj2012.0065.
Oliveira, J.C., Trabaquini, K., Epiphanio, J.C.N., Formaggio, A.R., Galvão, L.S., Adami, M., 2014. Analysis of agricultural intensification in a basin with remote sensing data. GIsci. Remote Sens. 51 (3), 253–268.
Pan, Z., Huang, J., Zhou, Q., Wang, L., Cheng, Y., Zhang, H., Blackburn, G.A., Yan, J., Liu, J., 2015. Mapping crop phenology using NDVI time-series derived from HJ-1A/B data. Int. J. Appl. Earth Obs. Geoinf. 34, 188–197.
Peña, M.A., Brenning, A., 2015. Assessing fruit-tree crop classification from Landsat-8 time series for the Maipo Valley, Chile. Remote Sens. Environ. 171, 234–244.
Rufin, P., Frantz, D., Ernst, S., Rabe, A., Griffiths, P., Özdoğan, M., Hostert, P., 2019. Mapping Cropping Practices on a National Scale Using Intra-Annual Landsat Time Series Binning. Remote Sens 11, 232.
Picoli, M., Camara, G., Sanches, I., Simões, R., Carvalho, A., Maciel, A., Almeida, C., 2017. Big earth observation time series analysis for monitoring Brazilian agriculture. Isprs J. Photogramm. Remote. Sens. 145, 328–339.
Ribeiro, J.F., Walter, B.M.T., 2008. As principais fitofisionomias do bioma cerrado. In: Sano, S.M., Almeida, S.P., Ribeiro, J.F. (Eds.), Cerrado: Ecologia E Flora. Embrapa Cerrados, Brasilia, DF, Brazil, pp. 153–212.
Rubinstein, R., Kroese, D., 2008. Simulation and the Monte Carlo Method, 3rd ed. Wiley – Interscience, Hoboken, NJ, USA, pp. 1–340.
Rudorff, B.F.T., Aguiar, D.A., Silva, W.F., Sugawara, L.M., Adami, M., Moreira, M.A., 2010. Studies on the rapid expansion of sugarcane for ethanol production in São Paulo State (Brazil) using landsat data. Remote Sens. (Basel) 2, 1057–1076.
Rufin, P., Muller, H., Pflugmacher, D., Hostert, P., 2015. Land use intensity trajectories on Amazonian pastures derived from Landsat time series. Int. J.Appl. Earth Obs. Geoinf 41, 1–10.
Sakamoto, T., Yokozawa, M., Toritani, T., Shibayama, M., Ishitsuka, N., Ohno, H., 2005. A crop phenology detection method using time-series MODIS data. Remote Sens. Environ. 96 (3-4), 366–374.
Sanches, I.D.A., Feitosa, R.Q., Diaz, P.M.A., Soares, M.D., Luiz, A.J.B., Schultz, B., Maurano, L.E.P., 2018. Campo Verde database: seeking to improve agricultural remote sensing of tropical areas. Ieee Geosci. Remote. Sens. Lett. 15 (3), 369–373.
Schmidt, M., Pringle, M., Devadas, R., Denham, R., Tindall, D., 2016. Framework for large-area mapping of past and present cropping activity using seasonal landsat images and time series metrics. Remote Sens. (Basel) 8, 312.
Schwieder, M., Leitão, P.J., Bustamante, M.M.C., Ferreira, L.G., Rabea, A., Hostert, P., 2016. Mapping Brazilian savanna vegetation gradients with Landsat time series. Int. J. Appl. Earth Obs. Geoinf. 52, 361–370.
Shapiro, D.E., 1999. The interpretation of diagnostic tests. Stat. Methods Med. Res. 8 (2), 113–134.
Spera, S., 2017. Agricultural intensification can preserve the brazilian cerrado: applying lessons from Mato Grosso and Goiás to Brazil's last agricultural frontier. J Trop Conser. Sci. 10 (1), 1–7.
Spera, S.A., Galford, G.L., Coe, M.T., Macedo, M.N., Mustard, J.F., 2016. Land use change affects water recycling in Brazil's last agricultural frontier. Glob. Chang. Biol. 22, 3405–3413.
Strassburg, B.B.N., Brooks, T., Feltran-Barbieri, R., Iribarrem, A., Crouzeilles, R., Loyola, R., Latawiec, A.E., Oliveira Filho, F.J.B., Scaramuzza, C.Ad.M., Scarano, F.R., Soares-Filho, B., Balmford, A., 2017. Moment of truth for the Cerrado hotspot. Nat. Ecol. Evol. 1 (99), 1–3.
U.S. Geological Survey, 2017. U.S. Geological Survey. Available online: (accessed on 10 April). https://landsat.usgs.gov.
Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sens. Environ. 185, 46–56.
Wardlow, B.D., Egbert, S.L., Kastens, J.H., 2007. Analysis of time-series MODIS 250m vegetation index data for crop classification in the US central Great Plains. Remote Sens. Environ. 108, 290–310.
Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., Woodcock, C.E., 2012. Opening the archive: how free data has enabled the science and monitoring promise of Landsat. Remote Sens. Environ. 118, 127–139.
Xu, H., 2005. A study on information extraction of water body with the modified normalized difference water index (MNDWI). Journal of Remote Sensing. 9 (5), 511–517.
Zheng, B., Myint, S.W., Thenkabail, P.S., Aggarwal, R.M., 2015. A support vector machine to identify irrigated crop types usingtime-series Landsat NDVI data. Int. J. Appl. Earth Obs. Geoinf. 34, 103–112.
Zhu, Z., Wang, S., Woodcock, C.E., 2015. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. Remote Sens. Environ. 159, 269–277.